

Collaborative Brain-Computer Interfaces in Rapid Image Presentation and Motion Pictures

Ana Matran-Fernandez

A thesis submitted for the degree of

Doctor of Philosophy

School of Computer Science and Electronic Engineering

University of Essex



March, 2017

Abstract

The last few years have seen an increase in brain-computer interface (BCI) research for the able-bodied population. One of these new branches involves collaborative BCIs (cBCIs), in which information from several users is combined to improve the performance of a BCI system.

This thesis is focused on cBCIs with the aim of increasing understanding of how they can be used to improve performance of single-user BCIs based on event-related potentials (ERPs). The objectives are: (1) to study and compare different methods of creating groups using exclusively electroencephalography (EEG) signals, (2) to develop a theoretical model to establish where the highest gains may be expected from creating groups, and (3) to analyse the information that can be extracted by merging signals from multiple users. For this, two scenarios involving real-world stimuli (images presented at high rates and movies) were studied.

The first scenario consisted of a visual search task in which images were presented at high frequencies. Three modes of combining EEG recordings from different users were tested to improve the detection of different ERPs, namely the P300 (associated with the presence of

events of interest) and the N2pc (associated with shifts of attention). We showed that the detection and localisation of targets can improve significantly when information from multiple viewers is combined.

In the second scenario, feature movies were introduced to study variations in ERPs in response to cuts through cBCI techniques. A distinct, previously unreported, ERP appears in relation to such cuts, the amplitude of which is not modulated by visual effects such as the low-level properties of the frames surrounding the discontinuity. However, significant variations that depended on the movie were found. We hypothesise that these techniques can be used to build on the attentional theory of cinematic continuity by providing an extra source of information: the brain.

Acknowledgements

Although there is only one name in the front cover of this thesis, the work presented in the following pages would not have been possible without the patience, guidance and help of my supervisor, Riccardo Poli.

Quiero agradecer especialmente a mis padres y a mi hermano su amor, ánimo y ganas de celebrar cada logro. To Ashwell, for his support and love throughout these years. No habría llegado aquí sin vosotros.

Many people have been involved in the writing of these pages, in different ways. I want to thank them for their help and patience while I was “too busy to talk”, or not in the mood to do so. A Laura, Clara, Yaiza, Pablo, Carlos... gracias por los abrazos a distancia y las reuniones anuales.

Big big thank you to my “Colchy” family: I wouldn’t have survived these years without the Beer-Cake Interface (BCI) group. To Meo for keeping me sane and for all the happiness and sweetness in the shape of cakes. To all the others for not allowing me to eat them alone (and all the other ways that you know you have of supporting me), most especially to Davide, Diego and Louis, but not forgetting the old and new “subscribers”: Javi, Florian and Miguel.

To the (real) BCI club for their stimulating meetings and the feedback on my work, and especially to Luca for allowing me to prioritise the writing of these pages in the last few months.

To my BrainStormers, with a specially BIG thank you to David and Hilary for believing in us and enduring the long sessions of live debugging, and for their hospitality.

To all my participants, who made this research possible by enduring my sometimes long and tedious experiments.

Last, but not least, to the School of CSEE in the University of Essex, for funding my PhD via a stipend grant, and to the EPSRC, whose grant EP/K004638/1 allowed me to travel to NASA JPL, where I collected much of the data presented in this thesis.

Contents

Abstract	i
Acknowledgements	iii
Contents	v
List of Figures	xi
List of Tables	xviii
List of Acronyms	xxiii
1 Introduction	1
1.1 Motivation	1
1.2 Research Objectives	4
1.3 Thesis Organisation	6
2 Literature Review	8
2.1 Introduction	8
2.2 Electroencephalography and Event-Related Potentials	10
2.2.1 The P300 Component	12

CONTENTS

2.2.2	The N2pc Component	14
2.3	Brain-Computer Interfaces	19
2.3.1	P300-Based BCIs	20
2.3.2	BCIs and the N2pc Component	22
2.3.3	Other Types of Single-User BCIs	23
2.3.4	Collaborative BCIs	26
2.4	Rapid Serial Visual Presentation	30
2.4.1	RSVP in Visual Search	31
2.4.2	RSVP and BCIs	32
2.4.3	Phenomena Associated with the RSVP Technique	38
2.4.4	Interim Discussion	41
2.5	Extreme RSVP: Motion Pictures	42
2.5.1	Psychological Effects of Structural Features in Movies	44
2.5.2	Psychophysiology of Structural Features	46
2.5.3	Movies and the Brain	51
2.5.4	Interim Discussion	53
2.6	Conclusions	54
3	Single and Collaborative BCI in Visual Search	56
3.1	Introduction	56
3.2	Methods	58
3.2.1	Data Acquisition	58
3.2.2	Experimental Design	60
3.2.3	Signal Processing and Feature Selection	63
3.2.4	Single-User Detection of Targets	63

3.2.5	Collaborative Classification	65
3.3	Results	67
3.3.1	Behavioural Results	68
3.3.2	ERP Analysis	68
3.3.3	Single-User BCI	71
3.3.4	Collaborative BCI	74
3.4	Discussion	77
3.5	Conclusions	80
4	Single and Collaborative BCI for Target Localisation	82
4.1	Introduction	82
4.2	Methods	84
4.2.1	Participants and Setup	84
4.2.2	Experimental Design	85
4.2.3	Feature Selection and Classification	86
4.3	Results	92
4.3.1	ERP Analysis	92
4.3.2	Single-User BCI	103
4.3.3	Handedness, Target Localisation and the N2pc	105
4.3.4	Collaborative Classification	110
4.4	Discussion	112
4.5	Conclusions	116
5	Concatenating Detection and Localisation Systems	118
5.1	Introduction	118
5.2	Methods	119

5.2.1	Participants and Setup	119
5.2.2	Feature Extraction	120
5.2.3	Sequential N–P System	121
5.2.4	Sequential P–N System	123
5.2.5	Sequential Regression System	126
5.2.6	Calculation of the Information Transfer Rate	132
5.3	Results	134
5.3.1	Sequential N–P System	134
5.3.2	Sequential P–N System	135
5.3.3	Sequential Regression System	138
5.3.4	Collaborative Sequential Regression System	140
5.4	Discussion	143
5.5	Conclusions	144
6	Improving cBCI Performance with Participant Selection	146
6.1	Introduction	146
6.2	Methods	148
6.2.1	Group Member Selection	148
6.2.2	Data	149
6.2.3	Comparisons with Other Systems	150
6.3	Results	151
6.3.1	Participant Selection in the Target Detection System . . .	151
6.3.2	Participant Selection for Target Localisation	157
6.3.3	Theoretical Model of the Participant Selection Method . .	167
6.4	Discussion	171

6.5	Conclusions	173
7	A Preliminary Study of Cuts in Movies Using Collaborative BCI	
	Techniques	175
7.1	Introduction	176
7.2	Methods	180
7.2.1	Data Acquisition	180
7.2.2	Stimuli	180
7.2.3	Signal Processing	182
7.2.4	Left/Right Attention Shifts	186
7.3	Results	189
7.3.1	Behavioural Results	189
7.3.2	ERP Analysis	190
7.3.3	Influence of Movie and Cut Characteristics on PCN	193
7.3.4	Shifts of Attention in Movies	195
7.3.5	Frequency Analysis	199
7.4	Discussion	205
7.5	Conclusions	208
8	Conclusions and Future Work	211
8.1	Summary of Contributions	211
8.2	Future Work	214
	References	218
A	List of Publications	248
A.1	Book Chapters	248

A.2	Journal Papers	248
A.3	Conference Papers	249
B	Comparisons Between the Collaborative BCI Methods	250
C	Participant Selection results	253
C.1	Group Member Selection on the Target Detection System	254
C.1.1	Median AUCs with Group Member Selection	254
C.1.2	Comparison with $\text{avg}(AUC_1, AUC_2, \dots, AUC_r)$	266
C.1.3	Comparison with $\text{max}(AUC_1, AUC_2, \dots, AUC_r)$	274
C.2	Group Member Selection on the Target Localisation System	282
C.2.1	Median AUCs with Group Member Selection	282
C.2.2	Comparison with $\text{avg}(AUC_1, AUC_2, \dots, AUC_r)$	294
C.2.3	Comparison with $\text{max}(AUC_1, AUC_2, \dots, AUC_r)$	302
D	Individual ERD/ERS Analysis in Response to Cuts	310
D.1	Individual ERD/ERS Topoplots	310
D.2	Grand Average Coherence Plots Preceding and Following Cuts	319

List of Figures

2.1	Different strategies to merge the brain activity from multiple users in collaborative BCIs.	27
3.1	Examples of target (a) and non-target (b) images used in the experiments.	61
3.2	Illustration of the protocol used in the experiment, for a presentation rate of 5 Hz. For clarity, target images are highlighted in this figure.	61
3.3	Target templates used in the RSVP paradigm. The template in (a) was used in all difficulty levels. Templates (b) and (c) only appeared in levels 3 and 5. Note that these are not the full images shown in the experiment, but fragments of them.	62
3.4	Electrodes used for P300 detection (i.e., for classification of target vs non-target).	64
3.5	(a) Stimulus-locked grand averages for T and NT trials at channel Pz and their differences and (b) scalp maps of the grand averages for T and NT at 297 ms and 515 ms after stimulus onset, for a presentation rate of 5 Hz.	70

LIST OF FIGURES

3.6	Distributions of mean AUCs across the cross-validation folds for single-user BCI across all levels of the experiment.	72
3.7	Scatterplot of the reported plane counts for each participant and difficulty level vs the corresponding AUC obtained in cross-validation for that participant and level of difficulty.	73
3.8	Median AUC values across participants (for size 1) and all possible combinations of participants plotted for every level when using the SC-cBCI method.	74
3.9	Median AUC values across participants (for size 1) and all possible combinations of participants plotted for every level when using the MC-cBCI method.	75
3.10	Median AUC values across participants (for size 1) and all possible combinations of participants plotted for every level when using the LDA-cBCI method.	75
4.1	Electrodes used for target localisation (based on the N2pc).	87
4.2	Electrodes used for evaluating the influence of the N2pc on target detection.	91
4.3	Contralateral and ipsilateral stimulus-locked grand averages at channels PO7 and PO8 and their difference (continuous line, “N2pc”) across lateral targets from the training set from one of the cross-validation folds.	93
4.4	Grand-averaged scalp distributions between 312 ms and 375 ms after the onset of LVF (top row) and RVF targets (bottom row) at a presentation rate of 5 Hz.	95

LIST OF FIGURES

4.5	Difference plot of the contralateral minus the ipsilateral grand-averages at channels PO7 and PO8 across all lateral targets from the training set for levels with only one type of target.	97
4.6	Difference plot of the contralateral minus the ipsilateral grand-averages at channels PO7 and PO8 for targets that are within 2–3, 4–5, 6–7, 8–9, 10–11 and 12–13 stimuli away from the previous target, and the grand average (labelled as “all”) across all lateral trials, for a presentation rate of 12 Hz.	100
4.7	Grand averaged contralateral minus ipsilateral differences at channels PO7 and PO8 for levels with one (i.e., levels 2 and 4) vs several (i.e., levels 3 and 5) types of targets at presentation speeds of (a) 6 Hz and (b) 10 Hz. Shaded areas represent time intervals where the two conditions are significantly different.	102
4.8	Distributions of mean AUCs across the cross-validation folds for single-user BCI across all levels of the experiment.	104
4.9	Contralateral minus ipsilateral stimulus-locked grand averages for LH and RH participants across lateral targets from the training set of one fold at the presentation rate of 5 Hz, and p values from a one-sided Mann-Whitney U test comparing both conditions. . .	106
4.10	Contralateral minus ipsilateral stimulus-locked grand averages for LH and RH participants across (a) LVF and (b) RVF targets from the training set of one fold at the presentation rate of 5 Hz, and p values from a one-sided Mann-Whitney U test comparing both conditions.	106

LIST OF FIGURES

4.11	Predicted x-coordinate for the target vs actual target position (in pixels) for all targets in the test set, using only LH participants. The regression line is also shown.	108
4.12	Probability density functions of the SVM normalised scores in the T vs NT classification for non-targets, lateral targets and central targets using combinations E_{28} (left) and E_{20} (right) electrodes, computed via R's Gaussian-kernel-based density estimator. Vertical lines represent the mean of each distribution.	110
4.13	Performance of the SC-cBCI in terms of the median AUC for each group size and level.	111
4.14	Performance of the MC-cBCI in terms of the median AUC for each group size and level.	112
4.15	Performance of the LDA-cBCI in terms of the median AUC for each group size and level.	113
5.1	Pipeline of the sequential N–P system.	121
5.2	Pipeline of the sequential P–N system.	124
5.3	Pipeline of the sequential regression system.	127
5.4	Confusion matrix from participant 3 at level 2 on the sequential P–N system.	137
5.5	Predicted vs real x-coordinate of targets, for the best performer at this task (participant 3, in level 2).	140
5.6	Predicted vs real x-coordinate of targets, for a group of size 7 at level 2.	142

LIST OF FIGURES

6.1	Distributions of mean AUCs for the cross-validation and test folds for the single-user target detection BCI across all levels of the experiment.	151
6.2	Surface interpolation of the AUC improvements obtained when using MC-cBCI for groups of size 3 over the AUC of the best member of the group.	157
6.3	Distributions of mean AUCs for the cross-validation and test folds for the single-user target localisation BCI across all levels of the experiment.	158
6.4	Probability density functions and histograms for the AUCs of the target (a) detection and (b) localisation systems, at difficulty level 2.	161
6.5	Probability density functions and histograms for the differences in pairs of AUCs (i.e., $ AUC_i - AUC_j $) from the target (a) detection and (b) localisation systems, at difficulty level 2.	163
6.6	Probability density functions and histograms for the average AUCs of pairs of users (i.e., $\text{avg}(AUC_i, AUC_j)$) from the target (a) detection and (b) localisation systems, at difficulty level 2, when only pairs of users with $\delta < 0.05$ are considered.	164
6.7	Surface interpolation of the percentage of AUC improvements over the AUC of the best member of the group using MC-cBCI for groups of size 2–9.	168
6.8	Expected gain of the joint AUC over the better participant of a pair when the distributions of scores for both classes are given by normally distributed random variables, $S_{i,c} \sim \mathcal{N}(\mu_c, \sigma_i^2)$, with $ \mu_{C_1=L} - \mu_{C_2=R} = 1$ and standard deviations $\sigma_i \in [0.3, 4]$	171

LIST OF FIGURES

6.9	Surface interpolation of the expected improvements in the AUC (in percentage) over the AUC of the best member of the group, for different group sizes, according to the theoretical model, with the same parameters used in Figure 6.8.	172
7.1	Grand average plot for cuts and non-cuts at electrodes (a) FCz, (b) CPz, (c) Pz, and (d) Oz.	191
7.2	Scalp distributions of voltage amplitudes 200 and 400 ms after a cut for three representative participants.	192
7.3	Median shot length as a function of median PCN.	194
7.4	Contralateral minus ipsilateral grand averages for left and right cuts across all participants.	198
7.5	Grand averaged spatial distribution of ERD/ERS after the occurrence of a cut at different time intervals (from top to bottom: [0,100], [200,300] and [400,500] ms, referred to cut onset), for theta, lower alpha, upper alpha, low beta and upper beta bands.	200
7.6	Spatial distribution of ERD/ERS after the occurrence of a cut for participant 003 at different time intervals (from top to bottom: [0,100], [200,300] and [400,500] ms, referred to cut onset), for theta, lower alpha, upper alpha, low beta and upper beta bands.	201
7.7	Average coherence between pairs of electrodes regions before (top row) and after (bottom row) the occurrence of a cut in different frequency bands.	202
7.8	Average coherence between pre- and post-cut conditions for pairs of electrodes.	204

LIST OF FIGURES

7.9 Average coherence between pre- and post- epochs for each electrode. [204](#)

List of Tables

2.1	Summary of BCI literature in which experts were used for RSVP tasks.	37
3.1	Parameters of the different levels of the experiment.	62
3.2	Average total plane counts reported by participants as a function of difficulty level. The total real count was 240 airplanes for all difficulty levels. For a quick reference, the presentation rate and number of different targets are also reported.	68
3.3	Summary of comparisons between the 3 modes for combining evidence to create collaborative BCIs: LDA-cBCI (“L”), MC-cBCI (“M”) and SC-cBCI (“S”). The symbol > is used to indicate that the method on the left is statistically significantly better than that of the right. The symbol = indicates that the methods on both sides are not significantly different. N/A values represent those cases in which there were not enough samples to perform the test.	77
4.1	Distribution of lateral targets across levels. The reported percentage of lateral targets is referred to the total number of targets within the level — there are a total of 240 targets per level.	85

LIST OF TABLES

4.2	Medians and Kruskal-Wallis p values for the peak amplitudes of the voltage differences between contralateral and ipsilateral channels for LVF and RVF targets at a presentation rate of 5 Hz.	95
4.3	Results of a one-sided Mann-Whitney U test with Bonferroni correction comparing peak amplitudes of the N2pc for different presentation rates (in levels with only one type of target). P values below 0.05 are statistically significant.	98
4.4	Median AUC values across all participants for each difficulty level and combination of electrodes, for T vs NT classification.	109
4.5	Summary of comparisons between the 3 modes of creating cBCIs: LDA-cBCI (“L”), MC-cBCI (“M”) and SC-cBCI (“S”). The symbol $>$ indicates that the method on the left is significantly better than the one of the right. The methods on both sides of symbol $=$ are not significantly different. N/A values represent those cases in which there were not enough samples to perform the test.	113
5.1	Median AUC values across all participants for each difficulty level for the sBCI (E_{20}) and the sequential N–P system, for T vs NT classification, and Bonferroni-corrected p values (last row).	134
5.2	Mean and standard deviation of the ITR for the sequential N–P system, across all participants and for each difficulty level.	135
5.3	Median AUC values across all participants for each difficulty level for the sBCI and the sequential P–N system, for left vs right target classification, and Bonferroni-corrected p values (last row).	136

LIST OF TABLES

5.4	Mean and standard deviation of the ITR for the sequential P–N system, across all participants and for each difficulty level.	138
5.5	Mean and standard deviation of the correlation coefficient, ρ , and the slope (β) of the regression line fitted to the test set across all difficulty levels.	139
5.6	Mean and standard deviation of the correlation coefficient between actual and predicted x-coordinate of targets for different group sizes.	141
5.7	Mean and standard deviation of the slope of the regression line fitted to the test set for different group sizes and difficulty levels. .	141
5.8	Ratio between the mean regression slope and mean correlation coefficients between the predicted and the actual x-coordinates of targets.	142
6.1	Median AUC values for the three types of cBCIs for target vs non-target classification for difficulty level 2, as a function of group size and the dissimilarity-index threshold δ	152
6.2	Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 2.	153

LIST OF TABLES

6.3	Median improvements over the average participant in the group when using collaborative BCIs for target detection at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group).	154
6.4	Median improvements over the best participant in the group when using collaborative BCIs for target detection at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.	155
6.5	Median AUC values for the three types of cBCIs for left vs right classification of targets for difficulty level 2, as a function of group size and the dissimilarity-index threshold δ	159
6.6	Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 2.	159

LIST OF TABLES

6.7	Median improvements over the average participant in the group when using collaborative BCIs for target average at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test group AUC vs average	165
6.8	Median improvements over the best participant in the group when using collaborative BCIs for target localisation at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.	166
7.1	Details of the movie clips used in the experiment.	181
7.2	Range of ratings and median ranking given by participants to each movie clip. Low numbers indicate higher preference. Rows have been sorted by the median of the rankings (last column).	190
7.3	P values from a two-sided Mann-Whitney test with Bonferroni correction for PCNs across different movies.	193
7.4	Average normalised scanpath salience (NSS) and overall concordance correlation (ρ_c) coefficient for each video clip.	197

LIST OF TABLES

BCI – Brain-Computer Interface

cBCI – Collaborative Brain-Computer Interface

ERP – Event-Related Potential

RSVP – Rapid Serial Visual Presentation

SNR – Signal-to-Noise Ratio

Chapter 1

Introduction

This chapter introduces the motivation, research objectives and organisation of this thesis.

1.1 Motivation

Brain-Computer Interfaces (BCIs) are devices that convert electroencephalography (EEG) signals from the brain into commands that allow a user to control another device without the help of the usual peripheral pathways. BCI technology was originally conceived as an augmentative device for people with severe physical and mobility impairments [Vidal, 1973]. Hence, research on this field initially focused on applications such as the matrix speller [Farwell & Donchin, 1988] that allowed completely locked-in users to communicate with their environment. However, BCI systems have been progressively refined throughout the years, giving rise to new applications for a more generic audience [Krepki *et al.*, 2007].

One of the most common BCI paradigms uses epochs (i.e., fragments of EEG recordings) that are referred to the onset of a stimulus [Luck, 2014]. In this paradigm, modulations in the EEG signals that are evoked by the stimulus — called Event-Related Potentials (ERPs) — can be detected and used to control an external device, e.g., a computer, wheelchair or robotic prosthesis.

EEG recordings are heavily contaminated by ongoing brain activity not related to the presentation of the stimuli, electromagnetic noise and muscular artifacts (mainly from the eyes and eyelids). Thus, in order to make the BCI system more robust, it is common to present the same stimulus (intercalated with other stimuli) several times and use averages of the neural response to each one. Of course, this implies sacrificing some of the speed of the system due to the increased number of stimuli presentations. However, the improvements in BCI performance that are obtained by this averaging technique can have a huge impact in user abandonment rates by disabled people [Huggins & Zeitlin, 2012; Huggins *et al.*, 2015; Riccio *et al.*, 2016].

Although BCIs have been around for a few decades, it has not been until recently that researchers and people in the industry have considered applications of this technology for non-disabled people. One of the new research paths that has emerged in the last decade concerns collaborative BCIs (cBCIs), whose operation is based on the aggregation of EEG signals from a group of users [Cecotti & Rivet, 2014; Wang & Jung, 2011; Yuan *et al.*, 2012].

The different ways in which the information from the users of a cBCI can be fused provide an alternative to the traditional averaging technique without sacrificing speed. Since each member of the group is presented with each stimulus once, the speed of the final system is the same that would be obtained in single-

trial, single-user BCIs (i.e., BCIs in which stimuli are only presented once, and no averaging is performed), while being as robust to noise as average-based BCIs. For instance, the EEG signals (or, when applicable, ERPs) from the group that forms the cBCI can be directly averaged, or the outputs of their individual BCIs can be weighed in order to obtain a final system decision. In this way, collaborative BCIs can dramatically increase BCI performance with respect to single-user BCI.

For this, cBCIs are suitable candidates to be moved out of the ideal research laboratory conditions and into the outside world, not only for active control of a device (i.e., active BCI), but also to measure people’s reactions to different events through a passive cBCI (e.g., in the movie industry).

Jobs such as those of intelligence analysts, who are required to monitor incoming streams of images looking for threats, can be performed by means of single-user BCIs [Gerson *et al.*, 2006; Healy *et al.*, 2010; Kruse & Makeig, 2007; Mathan *et al.*, 2006, 2008] We believe that this field could still benefit from using collaborative BCIs by improving target detection, specially if the decision fusion level includes a measure of the level of alertness of a BCI user, so the output of the system can be adjusted accordingly (and breaks can be suggested) [Myrden & Chau, 2016].

Decision making is another promising field where cBCIs can be applied. Studies have shown that decisions can be more accurate by measuring the confidence of individual users in their respective decisions [Valeriani *et al.*, 2015a]. This can be of paramount importance in defence and security, but also in other fields such as stock markets.

Collaborative BCIs, through the averaging of single trials across users, rather than across repetitions, can be a useful technique for looking at the neural re-

sponses that are evoked in single trials, when it is not possible or practical to repeat the same stimulus. For example, the reaction to the first presentation of an event/stimulus is often different from those produced in further repetitions. An example of this is found in movie watching, where the second or third repetition of the film, when the full story is known to the viewer, will change the impact of the film (and might be viewed as a waste of time by both researchers and volunteers) [Bridwell *et al.*, 2015]. However, it has been shown that there is a high inter-subject correlation (measured through fMRI imaging) in the brain’s response to watching movies [Hasson *et al.*, 2004, 2008]. Since EEG signals have a higher temporal resolution than the fMRI technique, they allow for better study of the brain’s reaction to specific events. It is possible to envision the cinemas of the future having BCI equipments to measure the reaction of their customers, or screening sessions in which the impact of different trailers for a same movie is measured through the neural responses of the viewers.

Last, but not least, advances in BCI research due to the interest of BCIs (and cBCIs) for able-bodied users can trigger advances in single user BCIs, therefore also helping those for whom these systems were initially developed. Examples of this include making experiments more engaging, reducing setup and training times, and advances in signal processing, both for improving classification performance and cross-session and/or cross-user transfer [Congedo *et al.*, 2013; Krauledat *et al.*, 2008; Liyanage *et al.*, 2013; Llera *et al.*, 2012; Vidaurre *et al.*, 2007].

However, for this transition from labs to real world applications to happen, more research is needed in order to understand the origin of the aforementioned performance improvements.

1.2 Research Objectives

This PhD thesis aims at investigating cBCIs in order to increase understanding of how they can be used to improve performance of current single-user ERP-based BCIs.

The main research questions that are addressed in this thesis are the following:

- I To study and compare different methods of creating collaborative BCIs using exclusively EEG signals.
- II To develop a theoretical model to establish where the highest gains may be expected from creating groups.
- III To analyse the information that can be extracted by merging signals from multiple users in single trials.

In order to answer the questions stated above, two scenarios were considered. The first one involved an active BCI applied to the automatic classification (by means of neural responses) of rapidly-presented aerial images. In this scenario, we want to improve the performance of single-user BCIs while keeping the groups as small as possible (assuming that the costs increase by employing a higher number of people to perform the task). Thus, it is an ideal situation to study Questions [I–II](#): how to fuse information from multiple users and under which circumstances are the gains in performance from cBCIs maximal.

As it was mentioned before, passive cBCIs can also be used to measure the collective reaction of a group of users to events that, for practical reasons, cannot be presented multiple times. Movies are sequences of related images in which discontinuities are included to convey a story in the limits of time imposed by

the medium of film. Thus, they can be reinterpreted in terms of the RSVP technique (plus sound) and, in this sense, it is interesting to measure the reaction of the viewers to those discontinuities (i.e., cuts). This is addressed in the second scenario considered, where feature movies were used to study Question III. In this case, the aim is not to improve single-user BCI systems, but to find a way to reduce the noise that is recorded with the ERP component of interest. Hence, Questions I–II, which are related to improving the performance of a system, will not be studied *per se* (i.e., the goal of this thesis is not to be able to build a system that detects when a cut has been presented), but some comments with respect to them will be done.

As a result of studying these questions, other aspects related to cBCIs, and to the specific paradigms used to perform study these systems are also explored.

1.3 Thesis Organisation

This thesis starts with a literature review that aims at giving the reader an overview of the main topics that are covered by the remaining of this work. This is presented in Chapter 2, where the reader is introduced to the concept of brain-computer interfaces and an overview of the basic aspects of these systems, together with a review of the main ERPs that will be explored throughout the rest of this work. This chapter also includes a review of the main aspects related to the presentation technique that is used throughout the following chapters, and of the main aspects related to the study of the neural activations that are produced while movie-watching.

The experimental part of the thesis starts in Chapter 3, which describes the

first experiment, and in which collaborative BCIs were used to detect the presence of objects of interest (i.e., targets) in rapidly-presented real-life aerial images. This work continues in Chapter 4, which attempts the more difficult task of locating a target in an image that is known to contain one.

The systems developed in these two chapters are combined to form one whole system in Chapter 5. Finally, before moving to the second scenario considered, Chapter 6 presents a method for selecting participants to form groups in cBCIs based on their individual performance. Thanks to this method, it was also possible to study the causes for the performance improvements achieved by cBCIs with respect to their single-user counterparts.

Chapter 7 presents the second scenario considered in this thesis, and studies the effects of cuts in motion pictures, both from an individual and a collaborative point of view.

The overall contributions and findings are detailed in Chapter 8, together with proposed avenues for future work.

Chapter 2

Literature Review

An introduction to brain-computer interfaces, event-related potentials and the main findings related to the scenarios considered in this thesis: the rapid serial visual presentation protocol and movie watching.

2.1 Introduction

The main aim of this PhD thesis is to investigate collaborative BCIs in order to increase understanding of how they can be used to improve performance of current single-user ERP-based BCIs. In order to fulfil this aim, we have selected three main research questions that will be addressed in the remainder of the thesis: (1) comparing different methods of creating cBCIs using only EEG signals, (2) developing a model that can explain when the highest gains in performance may be expected, and (3) analyse the information that can be extracted in single trials by merging EEG data from multiple individuals.

This will be done by studying two scenarios. The first one involves the au-

tomatic classification of satellite images presented at high speeds depending on whether or not an object of interest (i.e., target) is present in each of them; and the second one is a passive BCI (i.e., one in which the user is not trying to reach any specific objective, but rather the EEG system is used to monitor the user's reactions) that users wore while watching fragments of feature movies.

We hope that these will help us gain insights into the questions presented above and in Section 1.2, as well as other aspects related to cBCIs and the specific paradigms used in this PhD thesis. For this, in this chapter we review the current state of the art in BCIs (single and collaborative), in order to assess where work is needed and what can be expected from the scenarios and work carried out.

This chapter is organised as follows: first, we will begin by giving an overview into EEG and ERPs (Section 2.2), which are the basis of the work developed in this thesis. Specific ERPs that were expected to be found in the scenarios considered, such as the P300 and the N2pc are also described in detail in this section. We will then move to the field of EEG-based BCIs (Section 2.3, focusing mostly on those based on ERPs, both in the single-user and the collaborative cases. The next two sections will be devoted to the main aspects of the RSVP technique (Section 2.4) and the literature that has been devoted to study viewer's responses to watching movies (Section 2.5). From here, the chapter will conclude (Section 2.6) with some remarks about the information gained through this literature review and the gaps in the literature, that will be addressed in the following chapters.

2.2 Electroencephalography and Event-Related Potentials

Electroencephalography is based on the recording of electrical fields in the brain that are associated with the activity of large populations of neurons. The sum of each neuron's electrical fields creates a dipolar field that can be recorded on the scalp under two main conditions: (1) the neurons have to be synchronously active (i.e., they have to fire at the same time), and (2) they are placed in a geometrically favourable position (usually, with their axons running parallel to each other) [Luck, 2014].

Electricity generated in this way flows through areas that present lower resistance in the brain, which acts as a volume conductor. Moreover, if the electrodes are placed on the scalp, the skull and skin between the electrode and the brain of the subject introduce further distortion to the signals. For these reasons, the voltages recorded by an electrode do not necessarily represent all the neural activity that takes place within the brain (e.g., some dipoles may cancel each other), and they do not necessarily reflect the neural activity that occurs directly underneath that electrode. This is the basis of the *low spatial resolution* of electroencephalography [Luck, 2014; Rugg & Coles, 1995].

However, electricity flows at a speed near that of light, which allows for the recorded signals to reflect activity that is going on in that precise moment in the brain. This gives the EEG a *very high temporal resolution*. Taking advantage of this feature, after the first human EEG recordings by Berger [1929], this technique has been widely used in the field of neuroscience as a tool for the study of cognitive processes and behaviours that underlie the identification of a stimulus, attentional

processes, etc. [Rugg & Coles, 1995], e.g., through the study of event-related potentials.

Event-related potentials, or ERPs, are specific neural responses that appear in the EEG recordings and are associated with particular events (e.g., the presentation of a stimulus) [Luck, 2014; Pfurtscheller & Lopes da Silva, 1999]. Typically, recorded ERPs are a sum of signals of interest (called ERP components) and noise. Noise comes from different sources, e.g., ongoing brain activity, mains noise and muscular artefacts (mainly from the eyes and eyelids).

There are several ways of isolating an ERP component and reducing the noise. The most common technique to achieve this through averages. It is assumed that the ongoing brain activity recorded is a sum of the (constant) response to the stimulus plus (random) fluctuations and artefacts and noise which are not time-locked to the occurrence of the stimulus and which vary randomly across epochs. Thus, averaging epochs that belong to repetitions of the same stimulus will reduce the noise, enhancing the response to the stimulus, and the resulting waveform will largely represent the activity that is temporally fixed to the stimulus [Luck, 2014; Pfurtscheller & Lopes da Silva, 1999].

Once an ERP component has been isolated, it can be defined by a combination of their amplitude, polarity, latency (or range of latencies at which they can be expected) and scalp distribution (usually in relation to the international 10-20 system) [Kappenman & Luck, 2011; Rugg & Coles, 1995].

The following chapters will focus mainly on two specific ERP components: the P300 and the N2pc, which are reviewed below.

2.2.1 The P300 Component

2.2.1.1 Features of the P300 Component

The P300 component is one of the most studied ERP components, and amongst the most used paradigms for BCI [Reza *et al.*, 2012]. It is a positive peak (hence the “P”) that was first detected 300 ms after stimulus onset, although later experiments showed that it can appear with a latency of up to 900 ms. Its amplitude can reach 40 μV , which makes it relatively large (compared to other ERP components) and, as we will see below, possible to detect in single-trial BCIs with relatively high accuracies. The P300 has its maximum over the parietal/central area [Polich, 2011; Polich & Conroy, 2003; Rugg & Coles, 1995].

One of the reasons why this component is so widely used is the ease with which it can be elicited. The P300 is observed in any task that requires discrimination between different stimuli [Polich, 2007]. Thus, this form of ERP can be elicited by means of the oddball paradigm, in which a sequence of different objects (called “standard stimuli” and “targets”) is presented to the user. In the classic oddball paradigm, a primary task (distinction between targets and standard stimuli) is performed in parallel with a secondary task (e.g., counting the number of targets that appear) with the aim to keep the subject engaged, which in turn will enhance the amplitude of the P300, making it more easily distinguishable from the ongoing EEG activity [Polich, 2004a].

The 3-stimulus oddball, a variation of the standard oddball paradigm, includes a third type of stimuli: the distractors. These are non-target stimuli that differ from the standard non-targets in some form. The appearance of distractors (which do not necessarily have to be novel to the user) elicits a *novelty* P300, or

P3a component [Polich & Conroy, 2003], which differs from the *classic* P300 component (called P3b in this context) both in location and latency: it is recorded by electrodes placed on the frontal area of the scalp and has a shorter latency than the P3b.

2.2.1.2 Neurophysiology of the P300 Component

One of the theories that attempts to explain the neurophysiology and characteristics of the P300 and its two subcomponents, and the most widely adopted, is the context-updating theory. In it, Donchin [1981] states that the P3a subcomponent is linked to processes involved in the involuntary capture of attention by salient events, whereas the P3b reflects memory storage and serves as a link between stimulus characteristics and attention [Polich, 2004a,b, 2007, 2011; Polich & Conroy, 2003; Rugg & Coles, 1995].

2.2.1.3 Variations of the P300

The P300 is one of the most variable ERP components in terms of *latency* [Kappenman & Luck, 2011]. It has been reported to appear 300–900 ms after stimulus onset depending on several factors, such as the age and level of engagement from the individual, the difficulty of the task (i.e., the ease with which the target can be discriminated from the distractors and standard stimuli), and the type of task [Polich, 2007; Rugg & Coles, 1995].

Since much of the variability is due to the individual, it is not common to obtain grand averages across participants and these are usually performed across trials for a single individual to allow for inter-participant comparisons when studying the neurophysiology of the P300. Grand averages for an ERP show the earliest

onset and latest offset of the component [Luck, 2014], which can be quite different across participants: the times when the voltage is positive for some subjects could be negative for others and *vice versa*, resulting in a smaller average brainwave. Moreover, the level of engagement of the subject in the experiment will also result in variations of P300 latencies, causing a high intra-subject variability [Luck, 2014].

The *amplitude* of the P300 component depends on target probability and the target-to-target interval. P300s have a refractory period during which consecutive P300s, if at all elicited, will be of smaller amplitude. Hence, if targets appear very frequently, P300s elicited by consecutive targets will suffer from this [Polich, 2007]. However, the amplitude of the P300 will not depend on the target itself [Polich, 2007].

The amplitude of the P300 has also been linked to the attentional resources dedicated to the task that elicits it, so that it can be used as a physiological measure for testing resource allocation: when performing two tasks, the P300 elicited for each of them will have an amplitude that is a function of the relative attention that is given to that task [Näätänen, 1992; Polich, 2004a]. This result strengthens the resource allocation theory of attention: we all have a finite pool of resources that can be assigned to a task. The more resources are given to the primary assignment, the less will be left for others. Moreover, reductions in P300 amplitude for one task are correlated with a linear decrease of performance on the task [Donchin *et al.*, 1978; Isreal *et al.*, 1980]. For this reason, the P300 has also been proposed as a tool to measure user engagement [Donchin *et al.*, 1986].

2.2.2 The N2pc Component

2.2.2.1 Features of the N2pc Component

The N2pc component has been extensively studied in literature related to theories of selective attention. It is a negative component (hence the “N” prefix) that generally appears within 170–300 ms of stimulus onset, in the time range of the N2 component. The “pc” suffix denotes its location: posterior contralateral electrode sites (e.g., P7/8 and PO7/8 electrode sites from the 10-20 international system), meaning that it is elicited in electrode sites on the opposite side to the visual field where the target is found. The maximum amplitudes of the N2pc oscillate around 2–3 μV .

The N2pc is an asymmetric component which is best observed by computing the difference waves between the two brain hemispheres, so it is typically described in terms of the difference between the contralateral and the ipsilateral waves. For this, trial averages are performed separately for trials on which the target is ipsilateral or contralateral to a given electrode site [Hopfinger *et al.*, 2004]:

- The *ipsilateral* waveform includes the average of the left visual field (LVF) targets for the left hemisphere electrodes and the right visual field (RVF) targets for the right hemisphere electrodes.
- The *contralateral* waveform is calculated as the average of the RVF targets for the left hemisphere electrodes and the LVF targets for the right hemisphere electrodes.

2.2.2.2 Neurophysiology of the N2pc Component

Visual information about the environment enters the brain through the eyes and, more specifically, the fovea, which is the central area of the retina and provides about 2° of visual angle [Luck, 2012]. Before “foveating” to an object (i.e., moving the gaze so that the object falls in the middle of the fovea; overt attention), it is often necessary to perform a selective processing of the relevant features. This is usually done by means of covert attention (i.e., attending to a non-foveated object) [Luck, 2012].

The most extended theory of selective attention, called the feature-integration theory, was first proposed by Treisman [1969], who hypothesises the existence of specialized modules that automatically detect and code different sensory features (e.g., colours or shapes) in different feature maps. This is one of the first steps in covert attention. During visual search, the sensitivity will be increased for factors specified in the search template and the objects that comply with the pattern will have priority for further processing, producing a shift of covert spatial attention towards their location. Then, if the object still seems to match the search pattern, it may result in the item being “foveated” [Luck, 2012], which would be the first of several steps in overt attention.

The stages described above have been associated to different ERP components of visual perception. In particular, the N2pc component is related to the shift of covert spatial attention and reflects the focusing of attention on a potential target in visual search arrays, before foveating to it (for a review, see [Luck, 2012]).

The N2pc is often followed by the so-called Sustained Posterior Contralateral Negativity (SPCN), from which it is usually separated by a positive-going de-

flection (the origin and cause for which is not known). This component starts around 300–400 ms after stimulus onset and is found in visual search experiments in which the participants are asked to remember a given aspect of the target [Dell’Acqua *et al.*, 2006; Jolicoeur *et al.*, 2006a, 2008; Thiery *et al.*, 2016], but also in tasks that are not defined as memory tasks, but which are supposed to engage visual short term memory as a processing buffer (e.g., if participants are asked to make a choice afterwards) [Jolicoeur *et al.*, 2008].

In contrast to the N2pc, the amplitude of the SPCN is sensitive to the memory load, with bigger amplitudes associated with higher loads. Moreover, the SPCN persists for the duration of the retention interval [Jolicoeur *et al.*, 2008]. Hence, whereas the N2pc seems to indicate the orienting of attention towards a given hemifield (left or right), the SPCN is believed to indicate storage of the target in the visual short term memory [Dell’Acqua *et al.*, 2006; Jolicoeur *et al.*, 2006a, 2008].

Both the N2pc and the SPCN have been reported to be affected (by means of a decrease in amplitude and an increased latency) by the attentional blink [Dell’Acqua *et al.*, 2006; Jolicoeur *et al.*, 2006b; Zhang *et al.*, 2009] (see Section 2.4.3.1). However, even when the final responses given by participants are wrong, due to this effect, the N2pc has been observed in the EEG epochs [Dell’Acqua *et al.*, 2006; Woodman & Luck, 2003].

The SPCN is sometimes referred to as a Contralateral Delay Activity (CDA) [McCollough *et al.*, 2007; Töllner *et al.*, 2013] or Contralateral Negative Slow Wave (CNSW) [Klaver *et al.*, 1999]. However, it is still unclear whether these three components are actually the same or not [Jolicoeur *et al.*, 2008].

2.2.2.3 Variations of the N2pc

The list below summarises the main findings related to the N2pc and the factors that affect its *amplitude*:

- It is found when the search array contains at least one distracting item (any non-target stimulus) apart from the target. The closer the distractor, the bigger the N2pc amplitude will be [Luck & Hillyard, 1994b].
- It does not appear if the items surrounding the target provide information about it (e.g., when the experimental design requires two or more items to be compared) [Luck & Hillyard, 1994b].
- It is not elicited if there is only one item (regardless of whether it is a target or a distractor) in the search array [Luck & Hillyard, 1994b].
- It is elicited both in the presence of “pop out” targets and when discrimination is based on several conjunctive features, requiring sequential search [Eimer, 1996].
- It can be elicited not only by physical properties of stimuli, but also by semantic properties [Eimer, 1996].
- In upper vs lower visual field discrimination, the N2pc component is larger in the latter [Luck *et al.*, 1997].
- Distractors that closely resemble the target have been shown to elicit a smaller N2pc (which is not followed by a P300).
- The amplitude of the N2pc is relatively unaffected by target probability [Luck & Hillyard, 1994a].

-
- The amplitude of the N2pc is larger and it has a shorter latency when participants are given a target template rather than a target category [Nako *et al.*, 2014, 2015].
 - N2pc amplitude is smaller and has a longer latency when multiple targets are found in the search array [Nako *et al.*, 2014].
 - Searching repeatedly for a target in different search arrays enhances N2pc amplitude in the later trials [Nako *et al.*, 2015].

Even though the N2pc was first thought of as just a reflection of the attentional selection of target stimuli, the first three bullet points in the list above reflect its filtering function: when several items near the target compete for attention, a bigger effort to suppress the information from the distractors is required, so the N2pc has a larger amplitude (first bullet point). Moreover, this filtering is not necessary when the surrounding items give information that helps in the discrimination of the target (second bullet point) or when there are no competing items (third bullet point). Hence, the amplitude of the N2pc component would reflect the level of suppression of task-irrelevant or conflicting information.

If the feature that distinguishes the target from the distractors on the search array makes it “pop out” (e.g., if the distinctive feature is colour), it will be found faster (as evidenced by reaction times) than when finding the target requires sequential search. However, no statistical differences have been reported in the *latency* of the N2pc components elicited in these cases [Eimer & Kiss, 2007, 2008; Hickey *et al.*, 2006].

2.3 Brain-Computer Interfaces

A Brain-Computer Interface (BCI) converts signals generated by the brain into commands that serve as input to another device, e.g., a computer or prosthesis. In the most widely used type of BCIs, the electrical activity is recorded from the scalp of the subject via EEG and, after some type of signal processing and classification, a command is interpreted and sent to the external device. Since all the equipment is external to the user and no surgery is required for the functioning of the system, such BCIs are called non-invasive.

There are different paradigms that can be used to control a BCI by means of EEG recordings. As mentioned previously, one of them is through ERPs (e.g., P300-based BCIs). ERP-based BCIs have the advantage that they do not require the user to learn how to evoke the patterns in his/her brain in order to operate the system. However, phenomena such as habituation of the brain to the stimuli might, over time, reduce the performance of the BCI. Moreover, ERP-based BCIs require the user to be presented with stimuli, e.g., by means of a screen or through audio. In this review, we will focus on this type of BCIs (and, particularly, in those based on the P300 and the N2pc) as they are important for the remaining of the Thesis. For completeness, we have also included a review of other types of BCIs below (see Section [2.3.3](#)).

The first BCIs were developed with the aim of helping people with severe disabilities to communicate [Vidal, 1973]. These systems were progressively refined and current ERP-based systems are able to obtain relatively good performance in single trials (especially those based on the P300 component).

Due to these improvements, some forms of BCIs have recently started to be

explored with the able-bodied population in mind [Van Erp *et al.*, 2012; Wolpaw *et al.*, 2000], for instance focusing on the augmentation of human abilities, such as speed [Coffey *et al.*, 2010; Teplan, 2002] or the provision of a new means of control, for example, through passive BCIs that monitor the emotional state of the player in a game. Another focus of these new lines of investigation is on the use of brain activity recorded from multiple users simultaneously to increase the performance over single-user BCIs and non-BCI systems. These devices were first introduced as *hyperscanning* systems and used for passive applications, such as monitoring the brain activity [Babiloni & Astolfi, 2014], and as multi-mind (also called multi-brain or multi-user) BCIs for active control, e.g., to improve human performance in target detection [Wang & Jung, 2011], or play video games [Nijholt, 2015; Nijholt & Gürkök, 2013]. Despite the fact that this area of research has just recently appeared, as we will show below, the high number of papers published in this field and the broad range of suggested applications of collaborative BCIs have already shown its maturity and high potential.

2.3.1 P300-Based BCIs

One of the most common ways of controlling a non-invasive BCI is based on the use of the oddball paradigm. As we saw in Section 2.2.1, the occurrence of a target in a sequence of distractors triggers a P300 component. The ease with which this occurs has made P300-based BCIs very popular in research.

The first P300-based BCI was proposed and demonstrated by Farwell & Donchin [1988]. Their speller consisted on a 6x6 matrix that contained the alphabet and whose rows and columns were randomly intensified. The target in

each case was the letter that the user wanted to spell, eliciting a P300 when the row or column that contained that letter was intensified.

Matrix spellers and other BCIs based on the P300 usually rely on (involuntary) eye movements from the user [Acqualagna & Blankertz, 2011, 2013; Acqualagna *et al.*, 2010; Brunner *et al.*, 2010; Riccio *et al.*, 2012; Shishkin *et al.*, 2016; Treder *et al.*, 2011]. It has been argued that when a disabled person can still move his/her eyes, a speller or mouse based on eye-gaze is much faster and more accurate than current BCIs. Indeed, the performance of BCIs that rely on the P300, Steady-State Visual Evoked Potentials (SSVEPs; see Section 2.3.3) and/or any of its variations is drastically reduced when the user is a disabled person with impaired control of their gaze [Shishkin *et al.*, 2016]. Hence, Acqualagna *et al.* [2010] developed a new paradigm for the matrix speller based on the Rapid Serial Visual Protocol (RSVP) [Acqualagna & Blankertz, 2011, 2013; Riccio *et al.*, 2012; Treder *et al.*, 2011], where the different letters and symbols of the alphabet are presented at a rate at which eye saccades are almost completely suppressed (more on this on Section 2.4.2) [Potter & Levy, 1969]. The speller paradigm has further been explored in other types of BCIs, such as motor imagery BCIs (e.g., [Scherer & Muller, 2004]) and, more recently, steady-state visual evoked potentials (see [Cecotti, 2011] for a review).

Once the issue of communication was openly addressed for severely handicapped people and the usefulness of the matrix speller had been proved, the door was open for new applications that aimed at increasing the mobility of the users, e.g., by controlling a wheelchair with a BCI [Galán *et al.*, 2008], and in the form of assistive technology tools for interaction with the environment [Cincotti *et al.*, 2009].

One of the drawbacks of P300-based BCIs is the high variability of this component, which requires the system to be re-calibrated frequently if there is a change on the paradigm used to elicit the P300. In fact, some research has been done trying to adapt this component from one experimental setup to another with good results [Iturrate *et al.*, 2014]. This could help decrease training times, which would bring a significant improvement for all users and an increase in usability and acceptance [Huggins & Zeitlin, 2012; Huggins *et al.*, 2015; Riccio *et al.*, 2016].

Another main problem is the level of noise present in the EEG coming from both the brain of the user and his/her surroundings [Wolpaw & Wolpaw, 2012]. Because of this, the detection of a P300 by the classification system is typically based on the average of many signal repetitions of the same target events (i.e., the averages mentioned in Section 2.2). Averages need to be computed for every different stimulus and compared, so that the event with the highest averaged peak will be classified as the one selected by the user. The necessary repetition of the stimuli decreases the Information Transfer Rates (ITRs) and can make the system slow when high accuracy is required, as is the case in BCIs for disabled people, where high error rates would create frustration in the user. Hence, to avoid low information transfer rates, oddball sequences are usually presented at a fast pace, at the cost of eliciting deformed ERPs and, hence, lower accuracies on single-trial classification.

2.3.2 BCIs and the N2pc Component

Despite the well-known characteristics of the N2pc component and its appealing low variability, it has rarely been exploited to control a BCI.

Only a few studies have considered the N2pc (on its own or together with other ERPs) for controlling a BCI [Awni *et al.*, 2013; Blasco *et al.*, 2012]. Awni *et al.* [2013] displayed a circular array of white numbers around a fixation cross, following the experimental setup from Eimer & Kiss [2008]. Targets differed from the rest of the stimuli in their colour (red for targets, blue or green for distractors and white for non-targets). They found that N2pc components could help identify “pop out” targets accurately when averaging signals acquired over 3 repetitions of stimulus presentation. However, they also reported large variations in classification accuracy across individuals.

In a more practical experiment, Blasco *et al.* [2012] developed a P300 and N2pc-based BCI and applied it to three scenarios: web browsing (by controlling a keyboard and a mouse), a robotic arm and matrix-like communication tool. The last one was successfully used (with a classification accuracy of 80%) by a disabled participant after 6 repetitions of the stimuli.

With regards to the scope of this thesis, given the contralateral character of the N2pc component, we hypothesise that it could be beneficial in visual search scenarios for the delimitation of areas where the target is located within the whole image — something that has not been attempted before, to the best of my knowledge. Furthermore, given the small amplitude of this ERP component and its relatively low variability across individuals, it might be possible to obtain high performance improvements in cBCIs by averaging across subjects.

2.3.3 Other Types of Single-User BCIs

We have already mentioned that a big proportion of BCI systems rely on ERPs such as the P300 in order to work. These BCIs are called *synchronous*, since the user is given a cue and he/she is told when a command can be issued. Other form of synchronous BCIs include those based on Steady State Visual Evoked Potentials (SSVEPs). These are periodic evoked potentials that are generated in response to visual stimuli, such as a flickering or flashing light. If the frequency of the stimulus is higher than 6 Hz, it elicits an electrical response in the occipital area that can be used to control a BCI system. Typically, in an SSVEP-based BCI, multiple stimuli would flash at different frequencies, and the user selects one by focusing his/her gaze on it, so that the elicited response resonates at the same frequency as that of the stimulus [Sutter, 1992; Vialatte *et al.*, 2010], which makes the detection of the SSVEP relatively easy [Wang *et al.*, 2006].

One of the main advantages of evoked potentials, such as SSVEPs or ERPs is that users do not need to be trained to elicit them, as they are naturally evoked in the EEG in response to the stimuli. In this type of systems, only the classifier needs to be trained. An appealing aspect of SSVEP-based BCIs is that, in contrast to ERP-based BCIs, for which there may be much variability of latency and/or amplitude and are very sensitive to noise, SSVEP-based BCIs typically extract features in the frequency domain, and given the synchronisation of the signals in the occipital area (over the visual cortex), very specific and narrow ranges of frequencies are considered, which makes them more robust to noise [Wang *et al.*, 2006]. Moreover, they can be reliably detected using a low number of electrodes, which makes them a desirable candidate for daily use. The

limiting factor, then, is two-fold: (1) whether users can move their eyes to fixate on the intended stimuli [Morgan *et al.*, 1996], and (2) how long it takes for the EEG signal to synchronise with the external stimulus.

With regards to the first aspect, Allison *et al.* [2008] showed that, although gaze control might improve performance, it may not be necessary [Kelly *et al.*, 2005a,b; Lalor *et al.*, 2005], depending on the display and the task at hand.

Regarding the second issue, research has been conducted that relies on small modifications of the original SSVEP paradigm, such as code-modulated visual evoked potentials [Wei *et al.*, 2016], in which the stimuli no longer flash at a given constant frequency, and what changes is the phase of the stimulation between them. Recently, a non-invasive SSVEP-based BCI was developed that obtained the highest information transfer rate (including invasive BCIs) [Chen *et al.*, 2015], of 5.32 bits/second.

In contrast, there are other paradigms that require a conscious mental effort from the user. For example, this is the case of BCIs that operate based on cognitive tasks [Curran & Stokes, 2003], in which the BCI system learns the patterns of activation that emerge in the EEG as the user performs different mental tasks, such as imagining the movement of a limb (motor imagery BCIs) [Babiloni *et al.*, 2000; Penny *et al.*, 2000], performing mental calculations, or imagining a phone ringing [Dyson *et al.*, 2010]. In addition to not needing any form of external stimuli to be presented, these BCIs have the advantage of being fully independent of the user’s motor control capabilities, which makes them an ideal candidate for use by people with severe disabilities.

Finally, passive BCIs monitor the cognitive or affective state of the user and adapt the stimuli according to it [Myrden & Chau, 2016; Zander & Kothe, 2011;

Zander *et al.*, 2010]. For example, they can be used to detect the load of the user and suggest an appropriate time for the user to take a break, but also to give a measure of how reliable current EEG signals are (i.e., how similar to those that were used during training) [Myrden & Chau, 2016], an important aspect of BCI given the non-stationarity of the EEG, which can fluctuate significantly throughout a single session due to ongoing brain activity and other factors such as attention or fatigue [Cecotti & Ries, 2015; Marathe *et al.*, 2016].

2.3.4 Collaborative BCIs

As it was previously mentioned, BCIs have been traditionally developed as a form of assistive technology for people with severe disabilities. However, the advances in EEG technology and the development of relatively inexpensive headsets made possible the rise of BCI systems that can be used by able-bodied users as an extension of their capabilities [Cecotti & Rivet, 2014; Wang & Jung, 2011; Yuan *et al.*, 2012], i.e., as a new input channel.

A field that has arisen from these advances is that of collaborative BCIs, in which several users aim at jointly controlling one device simultaneously. In order to achieve this, the EEG signals from a group of users are merged or decoded together, so the final command that is sent to the device is derived from their collective intentions, rather than from a single user. It is worth noting that improvements derived from this field can in turn open the door to new clinical applications for single-user BCIs [Finke *et al.*, 2009]. For example, the current need for BCI systems to train a classifier at the beginning of each session in order to adapt to the user reduces the remaining time of practical use of the BCI [Fazli

et al., 2009a,b; Krauledat *et al.*, 2008; Mora *et al.*, 2015].

Model calibration (including feature selection and classifier optimisation and training) can be very time-consuming depending on factors such as the number of electrodes being recorded, so it is desirable for possible day-to-day use of this technology to reduce training times and increase transference, i.e., reduce the amount of data collected on the day to update a previously trained model (or classifier) while keeping a reasonable performance [Krauledat *et al.*, 2008; Manor & Geva, 2015; Marathe *et al.*, 2016]. Collaborative BCIs are likely to work toward decreasing training times and develop new methods for cross-session and cross-user transfer [Huang *et al.*, 2011; Sajda *et al.*, 2010] for example through the creation of user databases, so that the EEG signature of a new user can be compared to those from the database for a “plug and play” experience.

The aggregation of signals in collaborative BCIs can be done in different ways [Cecotti & Rivet, 2014; Li & Nam, 2016; Nijholt, 2015; Poli *et al.*, 2013a; Wang & Jung, 2011], as shown in Figure 2.1.

The simplest method consists of performing averages of the raw EEG signals across several users prior to their classification (*signal fusion* level). In this way, a unique classifier is used for all subjects and there is a reduction in the inherent levels of noise in the signals [Cecotti & Rivet, 2014; Cecotti *et al.*, 2014b; Jiang *et al.*, 2015; Kapeller *et al.*, 2014; Korczowski *et al.*, 2015; Poli *et al.*, 2013a]. However, this method may not be the most accurate, since the latency of time-locked ERPs (and, in particular, P300 components) varies both inter- and intra-subject depending on factors such as the level of attention, as was discussed in Section 2.2.1.

The second level of fusion illustrated in Figure 2.1 is the *feature level*. In

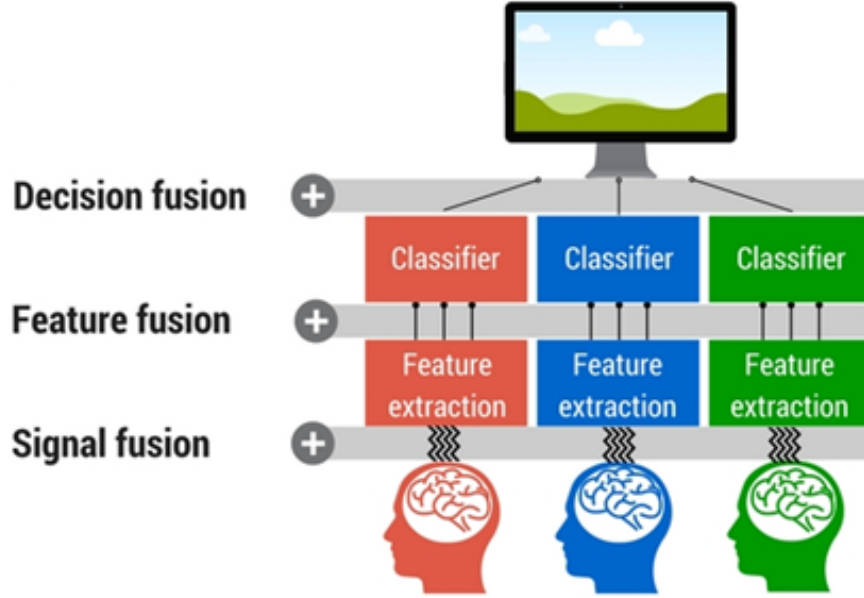


Figure 2.1: Different strategies to merge the brain activity from multiple users in collaborative BCIs.

this scenario, features are extracted from each user’s EEG. The fusion of features can be done by simple concatenation of them to form a unique feature vector for classifier training or any other combination [Eckstein *et al.*, 2012; Wang & Jung, 2011], so only one classifier is used (as in the signal level approach).

Finally, the information fusion can be done at the *decision level*, in which the EEG data from each participant is used to tailor one classifier specifically for him/her. In this case, a decision merging step needs to be implemented. At this level, we should emphasise the work from Cecotti & Rivet [2014]; Cecotti *et al.* [2014b], who studied different modes of combining the BCI decisions on a P300-based collaborative BCI and a SSVEP multi-brain BCI. Their strategies for merging the classifiers outputs included majority voting, averaging classifiers’ outputs, and maximum and minimum values. They found that averaging the classifiers’ outputs provided the best performance.

A considerable amount of work has been conducted to establish which level of fusion is optimal, obtaining consistent results across laboratories and applications. In particular, the two approaches that are often compared are the single-trial averages across participants (i.e., signal level) and fusion at the decision level (usually averaging classifiers' outputs to send a command). Since most of this work has been done based on different ERPs, given the inter-subject differences in latencies and amplitudes, it is not surprising that the best performance is obtained when information is merged at the decision level [Cecotti & Rivet, 2014; Cecotti *et al.*, 2014b; Wang & Jung, 2011].

Even though it is, in theory, possible to repeat trials for every participant also in the collaborative paradigm, it is expected that the classification will be done in single trials. Given the relatively big amplitude of the P300, this should not be a major problem for performance (especially if averaging across a sufficient number of users), but it could be a limiting factor for the detection of other ERP components, such as the N2pc.

Applications of collaborative BCIs include the simultaneous joint control of a single device. Wang & Jung [2011] showed that multiple users are able to generate and send a movement command to a prosthetic limb faster than single users. They also studied several methods of combining EEG recordings and reported an increase in the classification accuracy when the number of subjects was increased (they tested up to 20 individuals). Cecotti & Rivet [2014] obtained similar results on a P300-based matrix speller BCI in which up to 10 individuals were combined. Poli *et al.* [2013a] simulated an offline cBCI by combining signals from pairs of users to control a pointer on a screen by means of their BCI mouse [Citi *et al.*, 2008], and reported straighter trajectories than those achieved individually by

their participants.

Collaborative BCIs can also be found in the field of group decision making. In a scenario where it is not possible to average epochs over several trials (e.g., a person cannot be asked to make the same decision multiple times), aggregating the signals from a group of users in order to achieve a better outcome while reducing the level of noise has proven to be a useful technique [Poli *et al.*, 2013b]. Studies in this area of research tackle two problems: (1) whether group performance for decision making beats that of a single person, and (2) whether there is an optimal group size. Poli *et al.* [2013b] used a collaborative BCI to integrate the brain activity from up to seven subjects in a decision making task based on visual perception. By increasing the group size, errors were reduced. In further work, they extended their research (improving their results) by including non-EEG (i.e., behavioural) features, e.g., response times [Poli *et al.*, 2014].

Lastly, collaborative BCIs have been applied to visual search, using both search arrays [Valeriani *et al.*, 2015b] and naturalistic images [Valeriani *et al.*, 2015a]. The authors of this research found improvements in group performance on a target detection task when using a confidence estimator based on EEG signals and response times (hence using a hybrid BCI) to weigh each participant’s response on a trial-by-trial basis.

Collaborative BCIs can also be used as passive systems. In fact, the origin of cBCIs can be tracked down to 1965 if this type of systems is taken into account. Back then, they were known as *hyperscanning systems* [Babiloni & Astolfi, 2014]. Hyperscanning allowed researchers to discover that collaborative and competitive tasks have different effects on the connections in the brains of participants performing behavioural experiments (e.g., Astolfi *et al.* [2010]).

The concept of hyperscanning recently surfaced as passive multi-mind BCIs, examples of which can be found in the work of Hasson *et al.* [2004] and Hasson *et al.* [2008], in which the authors assessed the effect of feature films on brain activity during free movie watching. The main result of their work was to show that aspects such as movie content, editing and directing style have a direct impact on the level of control over the viewer’s brain activity. Later studies used passive multi-mind BCIs to show the high level of inter-subject correlation during natural vision [Bridwell *et al.*, 2015]. This discovery makes it possible to study the brain’s naive responses to stimuli by averaging signals across multiple users, hence increasing the low signal-to-noise ratio that is typical in EEG-based BCIs. Moreover, the high time resolution provided by EEG systems allows researchers to use this technique in multi-brain BCIs, for example, in those based on ERPs, which traditionally rely on multiple repetitions of a stimulus in single-user interfaces [Jiang *et al.*, 2015; Kapeller *et al.*, 2014; Korczowski *et al.*, 2015].

2.4 Rapid Serial Visual Presentation

In the rapid serial visual presentation (RSVP) technique, items are sequentially shown at a very fast rate on the same spatial location. Originally, this technique was designed for the study of reading and language processing by means of serial presentation of words or groups of words forming sentences [Forster, 1970].

The items displayed in RSVP are usually letters or images. Typically, the participants are given a particular search template before the stream starts. This description can be as specific or as generic as the researcher wants to make it (e.g., “animals” vs “male lions on a tree”). Participants are requested to either press a

button as soon as they see the target (typically enforcing either accuracy or speed, depending on the purpose of the particular study) during the presentation of the stream, or to observe a particular feature of the target and reply to a question after the block is finished (e.g., in which direction was the picture containing the target rotated) [Lawrence, 1971; Petrick, 1981; Potter, 1984].

2.4.1 RSVP in Visual Search

Typical tasks within RSVP experiments consist of matching a target named in advance (e.g., deciding whether a given letter or word had been shown in a stream of letters or words) [Petrick, 1981].

Potter & Levy [1969] studied recognition memory for pictures at presentation rates ranging between 0.125–2 seconds/picture. After a short stream showing the target pictures to look for, a new sequence in which distractors had been inserted was shown, and subjects had to verbally state at the end whether a given picture had been present in the film or not. They also studied eye movements and saccades during the presentation of the images: for sequences lasting between 2 and 4 seconds, eye movements were almost completely suppressed at rates greater than 4 pictures/second, whereas saccades were more common for lower presentation rates. This is relevant in the field of BCIs, because EEG systems are very sensitive to noise produced by eye movements and they should be avoided [Luck, 2014].

To date, there is no computer vision system capable of outperforming humans in visual search and recognition tasks. Such programs fail in the generalization of the description of the target, in a step that is straightforward for humans. The

same applies to situations in which we are not given a specific description of what type of object has to be found on images [Huber *et al.*, 2013]. For example, when performing broad-area search, intelligence analysts look for targets that fall in the category “threats”. One given object can mean very different things depending on its context, and a computer is not able to process these in the same way humans can. The search task can thus only be automatised to a very limited extent. Yet, manually analysing thousands of images is extremely tiring and time-consuming, so there is a need for software to help humans perform this task.

2.4.2 RSVP and BCIs

As described above, the RSVP technique consists of streams of stimuli (distractors) that appear sequentially on the same spatial location and, within each stream, participants are told to find a particular stimulus of interest (target). With this description in mind, it is not difficult to see similarities with the odd-ball paradigm.

By inserting a few target images amongst a large amount of distractors, or non-targets, BCI systems can detect a P300 component when an image that contains a target is presented within a stream of pictures [Gerson *et al.*, 2006; Healy *et al.*, 2010; Kruse & Makeig, 2007].

Moreover, since P300s are time-locked to the onset of the image that contained the target, the estimation of the temporal location of the image within the stream that can be obtained through the BCI has a lower variance than the one derived from user’s key presses [Huang *et al.*, 2011; Luck, 2014], and it can provide a continuous measure of confidence [Huang *et al.*, 2011].

Big organisations such as NASA and ESA have expressed their interest in using BCIs for the automatic classification of images by intelligence analysts [Healy *et al.*, 2010; Kruse & Makeig, 2007]. If successful, this could speed up the process of revising large amounts of pictures that are acquired by means of satellites and other sensors throughout the Earth.

Mathan *et al.* [2006] showed grayscale satellite images of a port in an RSVP task while recording EEG from volunteers in two experiments [Mathan *et al.*, 2006, 2008], at presentation rates of 10 and 20 pictures/second. Mathan *et al.* [2006] used a classifier trained with data from one participant to classify EEG epochs from the other, achieving high AUC values (0.84–0.85). In a second study, Mathan *et al.* [2008] presented images at a rate of 10 Hz to professional intelligence analysts. They showed that the BCI was able to speed up the target detection task without compromising detection accuracy (with respect to manual detection). This result was confirmed in their latest experiment [Huang *et al.*, 2011], in which they reported on an RSVP–BCI system (without any manual input) which was able to speed up the traditional broad area search paradigm by 5 times (measured in seconds/km²), a claim that has since been made repeatedly [Birisan & Beling, 2014; Marathe *et al.*, 2016; Touryan *et al.*, 2013]. Moreover, this work implemented an incremental learning (i.e., pre-trained classifiers are given additional samples to adapt their parameters, rather than starting the training from scratch each time) approach [Poggio & Cauwenberghs, 2001] in addition to cross-session generalisation, in which classifiers learnt using data from previous days (which was also successfully attempted by Manor & Geva [2015]).

Cecotti *et al.* have studied different aspects of the RSVP paradigm and the evoked P300 response to targets, such as target probability in the streams of

images [Cecotti *et al.*, 2011b], modality (i.e., audio vs visual RSVP, with and without key presses) [Cecotti *et al.*, 2011a], and the impact on attention of adding a second visual task in parallel to the RSVP stream [Cecotti *et al.*, 2012]. For instance, by varying target probability in a faces vs cars task (where participants had to press a key when they detected a face) they found that the behavioural performance and the amplitude and spatial distribution of the evoked potentials were significantly modulated by this parameter. They obtained the best classifier performance for target (face) detection at a target probability of 10% [Cecotti *et al.*, 2011b]. Moreover, they showed that it is possible to use a BCI for detection of targets through RSVP while multitasking by adding a secondary task (e.g., in addition to the RSVP target detection, they asked participants to press a button when they detected a green dot on a map that was shown next to the RSVP stream) [Cecotti *et al.*, 2011a, 2012; Marathe *et al.*, 2016]. Even though performance dropped for the easier task, it did not change significantly in the difficult one (in this case, the RSVP target detection task).

In addition to studying the effects of modifying aspects of the RSVP task, they studied ways of improving the performance of BCIs that used this paradigm to detect targets. One way in which they did this was by generating artificial trials by adding a small jitter and Gaussian noise to the stimulus onset reference in the EEG epochs, which serves a three-fold objective: (1) increase the size of the target class, hence combating the class imbalance and (2) training times, and (3) adding variability to the data, so that the classifier is invariant to small time shifts [Cecotti *et al.*, 2015]. In another experiment, they added a measure of the confidence from the user through the output of the classifier [Marathe *et al.*, 2015]. The effects of non-stationarity of the EEG data were also considered. Cecotti &

Ries [2015] showed that the way in which epochs are selected to train a classifier has an impact on its performance, due to factors such as tiredness or habituation effects [Marathe *et al.*, 2016]. Finally, another way to combat the changes in the EEG is by using active learning [Joshi *et al.*, 2012; Tong & Chang, 2001], an iterative semi-supervised technique which has been applied to situations in which there is abundance of data but obtaining the labels is expensive [Marathe *et al.*, 2016]. In each iteration of active learning, an active learner (i.e., a classifier) selects what are the data samples that are most informative, and it queries an oracle or expert for their labels, so that they can be added to the training set of the next iteration. In this way, the labelling effort is greatly reduced [Joshi *et al.*, 2012; Marathe *et al.*, 2016]. According to Marathe *et al.* [2016], the query to the expert can be done by providing feedback to the user and looking for error-related potentials (an ERP that occurs when the output of the BCI is not the one intended or expected by the user) [Marathe *et al.*, 2016], or through a behavioural response to that feedback. Perhaps, another interesting aspect of this work was the fact that, in some blocks, participants were presented with 500 ms long videos instead of static images as is traditional in image triage. The neural responses elicited by the short movies were more robust than those from the images [Marathe *et al.*, 2016].

Other research groups attempted the classification of images by combining EEG signals and computer vision [Kapoor & Shenoy, 2008; Manor *et al.*, 2016; Pohlmeier *et al.*, 2010, 2011; Sajda *et al.*, 2010; Ušćumlić *et al.*, 2013]. One of the most important findings from the iterative design developed by Ušćumlić *et al.* [2013] was the fact that their EEG classifiers were able to detect types of targets that were different from the training and the test sets, even though the

behavioural responses showed that some types of targets were more difficult to discriminate than others (a fact that could also be observed from the ERP averages, where some of the targets elicited larger P300 waveforms). In contrast, the approach taken by Manor *et al.* [2016] attempted joint detection of targets (buildings in satellite images, presented at 5 Hz and 10 Hz) by a multimodal neural network that fused information from the EEG and the image at the feature level. They tried two different approaches: a fully supervised system, in which the type of targets is known a priori, and a semi-supervised network, in which the type of target is unknown, so the computer vision component first trained an autoencoder on known non-target images and its output on target and non-target images was used as input features to the neural network together with the EEG response.

With respect to collaborative BCI approaches in the RSVP task, Yuan *et al.* [2012] performed a collaborative target detection task with groups of 3 people using visual evoked potentials in offline and online experiments¹. The BCI trained with data from the three participants was able to detect targets faster than the subjects' reaction times. Moreover, they reported an average increase of 11% in the accuracy of the group vs individuals (6% higher than the best individual).

Due to the speed at which the images are shown in RSVP experiments, participants do not usually have time to foveate to a target — that is, the target disappears before *overt attention* — the moving of the eyes in the direction of a salient item — can reach it. The reduction or absence of eye movements reported by Potter & Levy [1969] (at rates greater than 4 pictures/second) and Neider

¹When talking about offline/online BCIs we follow the traditional definition: *online* BCIs are those in which the brain signals of the participant are used at the time of collection to control the BCI. On the contrary, in an *offline* BCI data are recorded for posterior analysis, and the state of the interface is either static or manipulated by the experimenters.

et al. [2013] (for images shown for 150 ms or less) is particularly welcome in BCI systems based on RSVP stimulation, because EEG signals are severely distorted [Luck, 2014] by the artefacts produced by eye movements².

For this reason, RSVP-based BCIs pose a very attractive alternative for developing gaze-independent systems that are also suitable for severely locked-in people with no gaze control. Acqualagna *et al.* [2010] studied different presentation rates (83, 116 and 133 ms per item) in black and white vs colour (3 or 5 different colours) conditions [Acqualagna & Blankertz, 2011, 2013], and compared them to the traditional gaze-dependent speller from Farwell & Donchin [1988]. They found that the RSVP-speller paradigm was a suitable option for people with severely impaired oculomotor control. Their results were replicated by Treder *et al.* [2011], who performed a variation of the RSVP paradigm at a presentation rate of 5 Hz.

Before finalising this section and moving on to the limitations of the RSVP technique, it is important to note that research using the combined RSVP-BCI paradigm has been done not only on naïve participants, but also on experts in different fields, as shown on Table 2.1.

Moreover, a large amount of literature has compared behavioural responses with BCI performance and speed, and demonstrated that the combined RSVP-BCI paradigm is faster than traditional manual search and while being capable of maintaining the same levels of performance [Bigdely-Shamlo *et al.*, 2008; Birisan & Beling, 2014; Kapoor & Shenoy, 2008; Kapoor *et al.*, 2008; Parra *et al.*, 2008; Pohlmeier *et al.*, 2011; Poolman *et al.*, 2008; Sajda *et al.*, 2010; Touryan *et al.*,

²However, while eye *movements* during a trial produce performance-reducing artefacts in most BCIs, having one's eye gaze *fixated* on a target as opposed to, for instance, a fixation cross, significantly improves performance [Brunner *et al.*, 2010].

Table 2.1: Summary of BCI literature in which experts were used for RSVP tasks.

<i>Profession</i>	<i>Type of image</i>	<i>Relevant literature</i>
Intelligence analysts	Satellite imagery	Birisan & Beling [2014]; Healy <i>et al.</i> [2010]; Huang <i>et al.</i> [2011]; Kruse & Makeig [2007]; Mathan <i>et al.</i> [2007, 2008]
Transportation security officers	X-ray images of luggage	Trumbo <i>et al.</i> [2015]
Military	Urban landscape patrol simulation	Touryan <i>et al.</i> [2013]
Medical	Mammograms	Hope <i>et al.</i> [2013]

2013].

2.4.3 Phenomena Associated with the RSVP Technique

This section focuses on three well-known phenomena related to the RSVP technique: the Attentional Blink (AB), attentional awakening and the intertrial priming effect. Even though this review will not go very deep into these concepts, due to their relevance in the RSVP literature, it is important for the next sections to give the readers a clear idea of this phenomena in order to highlight the limitations of this technique.

2.4.3.1 The Attentional Blink

The attentional blink is a phenomenon that refers to a target being missed by participants when it is presented within a short time interval after the previous target in the same stream of stimuli [Chun & Potter, 1995; Cinel *et al.*, 2004; Einhäuser *et al.*, 2007; Kranczioch *et al.*, 2003].

The latency and duration of the attentional blink (i.e., the masking of the

second target) vary depending on the type of stimuli being presented [Einhäuser *et al.*, 2007; Most *et al.*, 2007].

A possible explanation of the AB is that after the detection of the first target, some resources will be allocated to the processing of that stimulus, temporally decreasing the capacity of the (limited) visual system [Di Lollo *et al.*, 2005; Kawahara *et al.*, 2006].

In order to avoid the attentional blink, Kruse & Makeig [2007] allowed a maximum of one target in every 51-picture stream. Furthermore, targets were placed within the central part of the image array to avoid crossover effects, since they reported that targets could not be detected in the last two images due to the “shock” produced by the end of sequence [Kruse & Makeig, 2007] (see Section 2.4.3.2).

Many of the studies described in the previous section have the same limitation with regard to the proportion of target vs non-target items, or on the way in which these are placed within a stream, e.g., allowing only one target within a stream of images [Mathan *et al.*, 2006, 2008]. On the other hand, Healy *et al.* [2010] constructed a more generic environment in which the proportion of targets vs non-targets was fixed at 10% — the typical target occurrence in the oddball paradigm, and an ideal ratio for eliciting large P300s [Johnson, 1986] —, and their results were not affected by this change. The same ratio was used by Cecotti *et al.* [2012, 2014a, 2015]; Marathe *et al.* [2015], although occasionally imposing restrictions on the inter-target interval to avoid attentional blinks [Marathe *et al.*, 2015].

When considering the effects of the AB in experiments and real life applications, the real expected number of targets should be taken into account. In broad-

area search, for example, the occurrence of targets may be very low (1% according to Mathan *et al.* [2008]), but targets could appear in images that show regions that are near each other. If neighbouring images are presented in close temporal proximity, as one would normally assume, this could produce attentional blinks and subsequent target misses. In this case, shuffling the images could solve the problem of the AB. On the other hand, the occurrence of images considered to be targets will be higher on medical data sets, a fact that should be taken into account in the labs, specially when deciding an appropriate presentation rate, since higher rates will make targets come closer in time.

Although it was recently shown that the AB can be decreased through training Choi *et al.* [2012], further developments in this area showed that the processing limits that are signalled by the occurrence of the AB cannot be eliminated [Tang *et al.*, 2014], and that the improvements achieved by Choi *et al.* [2012] may actually be related to the attentional awakening and intertrial priming effects (see below).

Lastly, even if a participant was able to detect the two targets, the refractory period of the P300 should be taken into account. In the case of two targets occurring very closely in time, the evoked P300 component (if at all present) will be of smaller magnitude, and might thus be misclassified by the system.

2.4.3.2 Attentional Awakening

As opposed to studies considered in Section 2.2.2, which looked into the *spatial* attentional filtering signalled by the N2pc, the RSVP technique is used to study the *temporal* modality of attentional filtering [Riccio *et al.*, 2013; Shapiro *et al.*, 1994], allowing the study of the temporal characteristics of neural information

processing [Chun & Potter, 1995].

The attentional awakening consists of a gradual increase in detection accuracy over time during an RSVP burst [Ariga & Yokosawa, 2008; Kranczioch & Dhinakaran, 2013]. This means that targets that are presented at the beginning of a burst are more likely to be missed than targets that occur at later time positions. According to Ariga & Yokosawa [2008], the attentional awakening represents the observer’s modulation of attention to adjust to the presentation rate in preparation for detecting the target in the sequence, since no significant variations of performance in target detection are found after the preparation is complete [Ariga & Yokosawa, 2008]. Kranczioch & Bryant [2011], however, argues that this phenomenon is not related to a variance in the focus of temporal attention, but rather to resource allocation to the task.

2.4.3.3 Intertrial Priming

The phenomenon by which repeating a target template improves the performance and speed of visual search is known as intertrial priming.

Intertrial priming was first proposed by Maljkovic & Nakayama [1994], and this effect has been observed in visual search scenarios regardless of whether the feature that remains constant for the target template is of the “pop out” type (e.g., colour) [Maljkovic & Nakayama, 1994; Yashar & Lamy, 2010a], requires a conjunctive search [Kristjánsson & Driver, 2008], or remains in the same spatial [Maljkovic & Nakayama, 1996] or temporal location [Kristjánsson & Campana, 2010; Kristjánsson *et al.*, 2010; Yashar & Lamy, 2010b].

Thus, this phenomenon is not only present in the RSVP paradigm, but might be related to the increased salience of primed features, thus facilitating the allo-

cation of attention to the processing of those features. Neurophysiology and neuropsychology studies have shown that this effect might be originated in the neural mechanisms that underlie visual search, including the attentional systems and regions in visual areas [Kristjánsson & Campana, 2010; Kristjánsson & Driver, 2008].

2.4.4 Interim Discussion

The RSVP technique, in conjunction with BCIs, has been studied for a number of years, which shows an interest in this technique to solve the real problem of rapidly classifying images. For this reason, we decided to use this joint RSVP–BCI approach as the first scenario for the study of collaborative BCIs presented in this PhD thesis.

Current computer vision systems are still not capable of identifying generic targets (e.g., those defined as “threats”) with the same ease as human beings do. We have shown in the literature review that the combination of a computer vision system together with physiological data from the human brain increases the detection of targets. However, this gain was only substantial when the computer vision was given the labels of the targets, and a drop of performance was observed when the same system did not know the type of target that it should look for [Manor *et al.*, 2016].

Only Yuan *et al.* [2012] attempted to create a collaborative BCI to detect targets, and reported a high gain in performance by using 3 participants. However, their experimental protocol (participants were asked to press a button as soon as they detected a target, which was presented at a random time after a

fixation cross) was much simpler than those typically attempted in a real RSVP paradigm. Regardless of this, their results, although possibly not fully applicable to the target detection aim of this thesis, seem promising.

This same observation can be made about Cecotti and collaborators' work, who investigated the effects of a wide number of parameters on a low-speed RSVP paradigm, predominantly at 2 Hz [Cecotti *et al.*, 2011a,b, 2012; Marathe *et al.*, 2016]. It has been shown [Yazdani *et al.*, 2010] that the ERPs evoked by different presentation rates vary greatly, and that they are attenuated for presentation rates above 10 Hz, so it is unclear how much of this work will be valid for higher presentation rates. However, as a starting point, we will take for valid his results on target probability [Cecotti *et al.*, 2011b] and keep the ratio of targets at 10%, as they suggested and has been done in many other works presented here.

Finally, as stated by Marathe *et al.* [2016], neural responses that are elicited by videos are more robust than those elicited by simple static images. Videos can be thought of as an RSVP presentation of highly correlated images shown at presentation rates of between 25–30 Hz (possibly with the addition of sound). We are interested in observing what occurs in the brain in response to a discontinuity in this “extreme form” of RSVP.

2.5 Extreme RSVP: Motion Pictures

We have suggested previously that a motion picture (or movie) can be thought of as an RSVP protocol in which images (i.e., frames) are shown at presentation rates above 25 Hz and which may contain sound. In addition to this, another feature of movies is that, most of the time, the images from consecutive frames

are very correlated with each other, giving the viewer a sensation of movement and continuity.

Movies tell a story, and in order to do so, different fragments (i.e., shots) are joined together, creating discontinuities in the stream, which represent an instantaneous displacement of one field of vision with another, and that may also be accompanied with a jump in time and/or space [Murch, 2001].

These discontinuities, called *cuts*, are introduced in movies to give viewers the impression of continuity by joining scenes or actions that are not continuous [Smith, 2012; Smith & Henderson, 2008b]. Even though cuts are not the only way to join different shots, they represent almost 99% of all transitions in contemporary movies [Cutting *et al.*, 2011], which is why they (and their effects on the viewer) have been studied in detail.

In his Attentional Theory of Cinematic Continuity (AToCC), Smith [2012] posits that what makes a cut continuous and acceptable for the viewer is proper cuing of the viewer across the cut. While watching a movie (but also in real life), the limitations of our working memory (which only allows us to maintain a maximum of four objects representations at a given time) [Kahneman *et al.*, 1992] imply that not everything we see is being attended to. Rather, our attention will be focused on a small and localized subset of the audiovisual features on the movie, and we will have some expectations of the next state of those features.

The decision of where we attend is a mixture of bottom-up (i.e., low level visual information, such as luminance and motion, which capture our attention involuntarily) and top-down factors (i.e., higher order cognitive factors, which present a voluntary decision to keep attention, for example, on the face of a character). In order to make a cut invisible, the filmmaker will have to direct the

gaze of the viewer by aligning his/her interest with compositional and low-level features on a scene to create attentional synchrony [Mital *et al.*, 2011; Smith & Henderson, 2008a; Smith & Mital, 2013] (more on this in the following sections). At this point, an additional stimulus, such as an onset of motion, or the voice from another character inside or outside the screen, will make the viewer turn his/her attention towards the source of change, producing a shift of *covert attention* if the source is off the screen, or a shift of *overt attention*, i.e., a saccade, if it is presented on the screen. If the cut is introduced at the point where attention is preparing to shift to another location, chances of viewers perceiving it are minimized, as long as continuity is preserved for their expectation of the features that were being attended to [Smith, 2012].

As we will show below, cuts and other forms of structural features (i.e., characteristics of movies that can be defined independently from the type of content, such as motion) influence attention and cognitive processes.

2.5.1 Psychological Effects of Structural Features in Movies

Before we look at the psychophysiological effects that movie-watching has on the viewer, we will give an overview of some of the main psychological aspects of movies. In particular, we will focus on three aspects of movies that are well known to affect the attention and emotion of their audience, and therefore, their engagement: motion, brightness and intensity [Cutting, 2016a,b].

As we have just shown, *cuts* in movies induce saccades and large eye movements on the viewer, and these eye movements are highly synchronised in time and space across viewers [Smith & Mital, 2013], a phenomenon known as *atten-*

tional synchrony [Mital *et al.*, 2011]. Since each saccade represents a perceptual enquiry (e.g., “What is she pointing at?”), it also elicits an expectation on the viewer about what the answer to that enquiry might be, and a shift of covert attention [Smith, 2012]. Hence, saccades drive attention [Mital *et al.*, 2011; Smith, 2012, 2013] and, for this reason, shot duration can be taken as a measure of the relative intensity of a movie [Cutting, 2016b; Wang & Cheong, 2006]. Short shots (and, correspondingly, a high number of cuts) reflect conflict and action, and can create anxiety on the viewer. Conversely, long shots are used in calm sections of the movie [Cutting, 2016a,b; Wang & Cheong, 2006].

In addition to the pace imposed by the frequency of cuts in a sequence, *motion* is also known to increase attention [Cutting, 2016a]. So much that comparing the amount of motion along old and contemporary movies shows significant differences, with the latter type having more motion contrast than the former [Cutting, 2014], perhaps in order to keep viewers engaged, a task which is more difficult given the lower thresholds of boredom in our current society [Cutting, 2005]. Across film genres, motion is highest in animated movies, and specially located on the center of the screen rather than the periphery [Cutting, 2014], possibly to keep children’s attention, but also in order to reduce production costs [Cutting, 2014].

Finally, *luminance* also drives attention [Cutting, 2014; Mital *et al.*, 2011], in addition to affecting mood and modulating emotion [Tarvainen *et al.*, 2015], with bright scenes being evocative of positive emotions and dark shots creating negative emotions [Cutting, 2016a,b; Troscianko *et al.*, 2012]. Recently, it was discovered that the “3/4-point” that is encouraged by screenplay manuals to be an extreme point on the emotional curve of a movie [Keating, 2011] actually

corresponds with the point of lowest brightness of the movie [Cutting, 2016a,b], possibly representing the moment at which the development of the story ends and the climax begins (i.e., the point on the story where events are most stacked against the protagonist) [Cutting, 2016a,b].

These three factors (and some others, such as sound and location of shots throughout a movie) were studied by Cutting [2016b], who quantified the number of cuts and the amount of motion and brightness (among other aspects of movie structure) throughout a number of movies distributed between 1935 and 2010. He performed a principal component analysis using these data and found that the first three principal components accounted for 75.7% of the variance in the data of movie structure. He named these components “editing” (43.4% of variance; largely represented by shot duration, scale and character introduction), “action” (20.7% of variance; represented by motion and location of action shots¹ during a movie) and “lightning” (11.6% of variance; represented by luminance), respectively. This result shows that these three aspects of a movie are indeed important and filmmakers are using them routinely to capture and manipulate viewers’ attention [Cutting, 2016b; Smith *et al.*, 2012].

Most of the work presented above refers to formal studies of movie features [Cutting, 2014, 2016a,b; Cutting & Candan, 2015], occasionally with eye trackers placed on viewers [Mital *et al.*, 2011; Smith, 2012; Smith & Mital, 2013]. The way in which different aspects of low level features (motion, brightness, etc.) affect viewers from an emotional point of view have also been tested using electrophysiological data, such as heart rate monitoring, galvanic skin responses, EEG

¹Action shots are defined as those “with beyond-normal physical activity”, not being restricted to fights, but also explosions, sports, and other extreme events [Cutting & Candan, 2015].

and functional Magnetic Resonance Imaging (fMRI), to measure the psychophysiological effects that movies have on viewers. This aspect of the literature is covered in the following section.

2.5.2 Psychophysiology of Structural Features

The two structural features that have been explored most exhaustively in the context of the psychophysiology of videos are motion and the reaction to discontinuities.

Both cuts and the onset of motion are known to elicit an orienting response (OR) [Lang *et al.*, 1993; Reeves & Thorson, 1986], which is an automatic response elicited by changes in the environment (either new, unknown elements, or learnt signals), whose purpose is to alert the person of changes that may require a response. This involves an increased attention towards the change that elicited it. The OR manifests physically in a person’s heart rate, skin conductance, and breathing patterns, amongst others. Thus, many of these aspects have been correlated with the onset or increase of motion in movies. Soleymani *et al.* [2008] reported changes in skin conductance and blood pressure associated with this. Similarly, variations in the pupillary responses [Ando *et al.*, 2002] and heart rate [Ando *et al.*, 2002; Cutting, 2016b] have been linked to motion onset. However, these changes may also be manifestations of the emotional response of the viewer to the content of the film, and thus, to their sense of presence (or immersion in the film) [Troschianko *et al.*, 2012].

Another effect of the orienting response is a decrease of power in the alpha band of the EEG, also termed *alpha blocking*, which is associated with an increase

of attention. Thus, the effects of cuts on viewers have also been studied through electroencephalography, specially associated with memory of content after viewers watched a number of television commercials [Appel *et al.*, 1979; Lang *et al.*, 1993; Reeves & Thorson, 1986; Reeves *et al.*, 1985; Rothschild *et al.*, 1986; Smith & Gevins, 2004], showing that indeed alpha blocking occurred after a cut.

This effect was first reported by Reeves *et al.* [1985], who also showed a co-variation between movement occurrences and alpha blocking. In their work and others, these drops occur within the first 500 ms after the cut, and alpha power increases again around 2 s after the cut. These drops of power in the alpha band have repeatedly been related to an increase of attention (e.g. [Lang *et al.*, 1999, 2000; Rothschild *et al.*, 1986; Smith & Gevins, 2004]). However, the magnitude of the drop and the rate of recovery may be related to interest in the content and its novelty (presumably because less resources are needed with successive repetitions of a stimulus) [Rothschild *et al.*, 1986].

The time to recovery of alpha may be taken as a measure of how well attention is held by the content. For example, not all scene changes are strong enough to elicit alpha blocking (i.e., they are not capable of gaining viewer attention), whereas other scenes, despite gaining attention (as evidenced by the sharp drop of power in alpha) may not hold it (evidenced by rapid recovery to previous levels). This may be connected with the concept of *edit blindness*, a phenomenon by which some cuts go completely unnoticed by viewers [Smith & Henderson, 2008b]. The probability of detecting a cut seems to vary depending on factors such as the type of cut and the amount and direction of motion and audio before and after it [Smith, 2012; Smith & Yvonne Martin-Portugues Santacreu, 2016]. For instance, Smith & Henderson [2008b] showed that match-action cuts (those

in which there is an onset of movement before the cut that continues after it, from a different point of view) are much more likely to be missed in a cut detection task using Hollywood films.

We have previously shown that cuts are also capable of affecting and directing attention, but not all cuts do so in the same way. One possible taxonomy of cuts divides them into two categories according to the content of the two shots that they join: context changes or *unrelated cuts* are those in which the scene after the cut is not directly related to that preceding it, an effect that is equivalent to changing the channel on the television. In contrast, *related cuts* are those in which the shots before and after the cut are related.

Lang *et al.* [1999] related responses to questionnaires about content retention of television commercials with power in the alpha band, and showed that these two types of cuts affect attention in different ways. Related cuts provide less new information and thus impose a lower cognitive load than unrelated cuts. Moreover, related cuts facilitate retention of new information, a skill that follows an inverted U-shape profile when plotted against shot length, showing that introducing more related edits can help retain clip content provided that they are not overused [Lang *et al.*, 1999, 2000].

However, these studies focused on the alpha band as the whole 8–12 Hz frequency range on occipital scalp regions. Contemporary works may prefer to study two sub-bands of the alpha range: the low alpha band (8–10 Hz, most predominant over prefrontal and parietal areas) concerns working memory processes and episodic memory encoding; in contrast, the upper alpha band (10–12 Hz, found in occipital and occipito-parietal areas) is most responsive to visual stimulation.

Smith & Gevins [2004] found a high correlation between the number of cuts

in commercials and the ratings of interest, adding to the corpus of knowledge that indicates that cuts help keep the viewer engaged. In the same study, they decomposed the alpha rhythm into the upper and low alpha bands and different scalp regions through a principal component analysis, and correlated the first three components (which explained 94% of the variance in the data) with the frequency of scene changes (i.e., shot length), subjective measurements of interest, and recall probability of the advertisement’s content. The first component (43% of variance) included power in the low and upper alpha bands in posterior electrodes, and was significantly correlated with scene change frequency. Thus, it was likely to reflect an automatic orienting of attention in response to a cut. The second component (26% of variance) was composed by power in the upper alpha in frontal electrodes, and significantly correlated with subjective interest in the commercials. Finally, the third component (25%) comprised lower alpha activation in frontal electrodes, and was significantly correlated with recall probability. Although this work involved 30-second long television commercials and not movies, the approach followed was the same as in Cutting [2016b]. It would be interesting to see how these components relate to the “editing” (which also explained 43% of the variance in the movie data, as did the first component of this work), “action” and “lightning” factors found in his work.

Anderson *et al.* [2006] performed, perhaps with the exception of Zacks *et al.* [2010] (see below), the only comparison between related and unrelated cuts using fMRI. In addition to the high similarity of neural activation found in the works of Hasson and collaborators (see below), they found a difference between coherent and incoherent shots in the activation with Brodmann area 31, which was only activated in the former, perhaps indicating the processing of information to create

meaning.

Cuts are very sharp in the time domain, whereas changes in the alpha band require some time, and fMRI techniques do not have the high time resolution that the EEG provides. For this reason, ERPs, which are time-locked to the stimuli and occur much faster than changes in frequency, have also been used to study the brain response to cuts [Francuz & Zabielska-Mendyk, 2013]. In this area of research, the occurrence of cuts has been associated with Slow Cortical Potentials (SCPs), which are also used as an index of the orienting response, and other ERPs [Francuz & Zabielska-Mendyk, 2013], such as the Slow Negative Wave (SNW) 1 (a frontal ERP with a peak around 448–648 ms), the Slow Positive Wave (SPW, a parietal component in the same range of latency as the SNW1), and the SNW2 (a parietal ERP found around 648–1800 ms after stimulus onset). Francuz & Zabielska-Mendyk [2013] studied the brain’s response to related vs unrelated cuts, and found that unrelated cuts produce more negative SNW1 and more positive SPW responses than those produced by related cuts [Francuz & Zabielska-Mendyk, 2013]. Although the authors also expected to find a P300-like ERP to signal attention shifts in response to cuts, they did not find it.

Another ERP that has been detected while presenting video stimuli is the N400 [Reid & Striano, 2008; Sitnikova *et al.*, 2003, 2008], a negative peak with a latency of around 400 ms after the presentation of the stimulus which is linked to meaning processing and has a bigger amplitude when a new element is incongruent with the context built by previous stimuli [Kutas & Federmeier, 2011] (for a review, see Lau *et al.* [2008]). In movie stimuli in which congruent or incongruent information appeared following a cut, an N400 can be detected over frontal, central and parietal regions [Reid & Striano, 2008]. The N400 was followed by

a late large positive potential in posterior regions of the brain in the sequences with an incongruity [Sitnikova *et al.*, 2003, 2008], perhaps an indication that this unexpected turn of events triggers other cognitive processes in addition to the ones that mediate pure semantic integration.

While these results indicate that cuts produce characteristic ERPs, these have always been studied in either averages or grand averages (i.e., after averaging hundreds of individual responses across several participants). Moreover, to the best of my knowledge, no ERP analysis of the neural response has been performed using feature movies: the work surveyed until now involved either commercial messages, audiovisual content created specifically for the experiment, or short fragments from television shows.

2.5.3 Movies and the Brain

Although a considerable amount of work has been devoted to understanding the psychological effects of different aspects of movies, and this work has been performed using real Hollywood films, it is remarkable how few studies using brain imaging techniques has been done in this area.

In one of the first fMRI studies concerning feature movies, Bartels & Zeki [2004] found that selected regions of the brain were activated in response to sequences containing color, faces, language, and people, corroborating previous fMRI studies that resorted to simpler, highly fabricated stimuli with the same features.

Another work published in the same year assessed the effect of feature films on brain activity during free movie watching [Hasson *et al.*, 2004]. This experiment,

in which participants were shown a 30-minute long fragment of a Hollywood movie, explored the degree to which filmmaker and editor’s manipulations impact the level of control over the viewer’s brain activity. They reported that a remarkable 45% of activation of the neocortex (including regions involved in vision, language processing, emotion and multi-sensory integration) were very similar throughout the duration of the movie across viewers and, and that this similarity increased with the degree of attentional synchrony [Anderson *et al.*, 2006; Hasson *et al.*, 2008].

More recently, this high level of inter-subject correlation during natural vision has also been shown to occur through EEG [Bridwell *et al.*, 2015]. This discovery makes it possible to study the brain’s naïve responses to stimuli by averaging signals across multiple users, hence increasing the low signal-to-noise ratio that is typical in EEG-based BCIs. Moreover, the high time resolution provided by EEG systems allows researchers to use this technique to look for viewer’s cognitive responses to specific events in the movies.

Zacks *et al.* [2010] had participants watch a short film while being inside an fMRI scan, and then asked them, by watching the movie two more times, to segment it into coarse and fine event segmentation. The events highlighted by the participants corresponded with transient changes in the fMRI signal, again highly synchronised across participants. Although no explicit mention of cuts is done in this work, it is probably a safe assumption that the segments pointed out by the participants were aligned with cuts. The changes in fMRI recordings were larger for fine than coarse changes, which, if the former were related cuts and the latter were unrelated cuts, would contradict all the evidence presented until now. However, two caviats should be pointed out here: first, we do not know

whether the segmentation actually coincided with cuts and, if so, what types of cuts, and second, as the authors point out, these results could be a result of the time scale. Moreover, the small time resolution of the fMRI technique might limit these results when compared with those from EEG analyses presented in previous sections.

As mentioned in previous sections, due to the noise affecting EEG recordings, in BCI it is common to perform averages across a small set of repetitions of the same stimuli. This has the benefit of increasing the signal-to-noise ratio of the signals, leading to better BCI performance. However, this technique cannot be used in the movie watching scenario, because the reaction to the first presentation of an event is different from those produced in further repetitions of the event [Dorr *et al.*, 2010; Nittono, 2008]. For example, the magnitude and duration of alpha blocking after a cut changes with repetitions of the same content [Rothschild *et al.*, 1986], and this habituation effect is also observed in the beta range (14–36 Hz) [Krugman, 1971]. For example, Bridwell *et al.* [2015] showed that it is possible to know whether a person had previously watched a movie by means of EEG recordings thanks to the high inter-subject correlations.

2.5.4 Interim Discussion

Movies can be seen as sequences of rapidly presented images that are very correlated to each other and which are usually accompanied by an audio track. However, they are known to elicit more robust neural responses than the RSVP protocol alone. For this reason, we decided to use the neural response to cuts in movies by means of collaborative BCIs.

Thanks to the effect of attentional synchrony, i.e., the high synchronisation of attention across viewers while watching dynamic images (e.g., movies) both in time and space, which has been measured through eye gaze and brain imaging techniques, we can average EEG signals across users to obtain a higher signal-to-noise ratio that might help develop an understanding of the neural response to cuts in a cut-by-cut basis.

We chose to focus on cuts because they represent almost 99% of all transitions found in contemporary movies, and they have been known to affect attention and the emotional state of the viewer (e.g., short shot lengths can create anxiety, and long shots create a calm sensation) for a few decades.

Motion and brightness can be used by film editors to try to control the attention of the viewer in order to create attentional synchrony so that cuts are not perceived as a disruption by the audience. For this, we will also study the specific influence of these two factors on the amplitude of ERPs.

Finally, as we have just shown, cuts in movies induce saccades and large eye movements on the viewer. Since each saccade represents a perceptual enquiry, it also elicits an expectation on the viewer about what the answer to that enquiry might be, and shifts of covert attention may arise as a result when the answer is not shown on the screen. Hence, we will try to determine whether the N2pc appears as a result of such perceptual enquiries after the occurrence of cuts.

2.6 Conclusions

The main goal of this PhD thesis is to increase understanding of cBCIs and to show how some of the collaborative BCI techniques can help advance research

when only single trials from each individual are available.

For this, we have envisioned two scenarios of real practical value. The first one, which uses the RSVP technique for triage, has been validated in single-user cases, but improvements can still be done. Even though a current trend is to merge physiological data (EEG) and computer vision algorithms, it is a common belief that the human vision is still ahead from computer vision, so we will use a collaborative BCI approach to show how, and under which circumstances, improvements can be made in this area of research.

The second scenario will look at Hollywood movies (which we consider to be a special case of the RSVP presentation in which images are highly correlated and presented together with sound) and study the neural response to cuts in a cut-by-cut analysis never attempted before, to the best of our knowledge. The reasons behind this gap in the literature might be due to EEG signals being noisy, plus the fact that showing the same videos multiple times changes the neural response to them.

However, recent developments in research on the phenomenon known as attentional synchrony, which has now been demonstrated using brain imaging techniques, may indicate that it is possible to average the brain response to a given trial across individuals rather than averaging multiple repetitions of the same stimulus for a given participant, a technique that has been used in collaborative BCIs over the last few years.

Chapter 3

Single and Collaborative BCI in Visual Search

This chapter presents a BCI based on the RSVP-oddball paradigm with the aim of classifying images with no other input from the users than their EEG signals. Different presentation rates were used and the results were analysed for single and collaborative BCIs using different methods to merge the evidence from the individuals in a group.

3.1 Introduction

As we showed in the literature review (see Section [2.4](#)), BCIs can be used for target detection in triage systems through the elicitation and detection of specific event-related potentials (i.e., the P300) that are evoked in the presence of targets in streams of pictures.

Different aspects from this paradigm, such as the optimal ratio of target vs

non-target images have been previously studied [Cecotti *et al.*, 2011b]. However, to the best of our knowledge, no real RSVP study has considered the differences in presentation rate on the detection of the P300 ERP. Perhaps, the most relevant work on this topic is that of Yazdani *et al.* [2010], who investigated the amplitude of the P300 by varying the presentation rate while keeping the proportion of targets fixed at 10% (the ideal ratio according to Cecotti *et al.* [2011b]). However, in their work images did not follow each other as in the traditional RSVP technique, but a neutral background screen was presented after each image for the same duration as the target image. Images were presented in 40-second long bursts at increasing rates from 500 ms/image (therefore followed by 500 ms of background) to 50 ms/image (followed by 50 ms of background), with the smallest presentation rate obtaining the highest P300 peak amplitude. Even with the effect of presentation rate on P300 amplitude, their classifiers were able to successfully discern between target and non-target images until a rate of 100 ms/image.

This chapter begins by establishing a baseline on single-user performance keeping the recommended target vs non-target ratio at 10% and using different presentation rates. Then, we will use the signals collected from multiple participants to assess the improvements that can be achieved by different ways of merging EEG evidence from several viewers (or users).

We should note that our approach is different from that used in many other RSVP-based BCI systems for target detection. In most cases, participants are required to press a key when they see a target (sometimes called the behavioural task) [Cecotti *et al.*, 2015; Gerson *et al.*, 2005, 2006; Marathe *et al.*, 2014, 2015; Parra *et al.*, 2008; Touryan *et al.*, 2014]. Then, only the trials in which the target was correctly identified by the participant (e.g., by including only those trials that

were followed by a key press within a pre-defined time period) are used for plotting grand averages and training a classifier with clean data. However, this method has several drawbacks [Sajda *et al.*, 2010], including *artefact contamination in the EEG resulting from the key press*.

This chapter is organised as follows: we start by describing the experimental setup, data collection and processing that are common for both the single-user and the collaborative BCIs (Section 3.2), and then present the results in Section 3.3. Finally, these results are discussed in Section 3.4 and the basis for the work presented in the following chapters is set in Section 3.5.

3.2 Methods

This section describes the data acquisition process, the experimental protocol followed and the preprocessing and classification algorithms that were used on the data in order to create the single-user and collaborative BCIs.

3.2.1 Data Acquisition

Data were gathered from 11 participants (mean age \pm standard deviation = 24.3 ± 3.9 years old, 4 females, 5 left-handed). All had normal or corrected-to-normal vision and reported no family or personal history of epilepsy. They all signed the consent form approved by the Ethics Committee of the University of Essex.

Participants were seated at approximately 80 cm from an LCD screen where the stimuli were presented.

EEG data were acquired by means of a BioSemi ActiveTwo system with 64 electrodes mounted in a standard electrode cap following the international 10-

20 system, including electrodes on the earlobes of the volunteers. EEG was referenced to the mean of the electrodes placed on the earlobes. The initial sampling rate was 2048 Hz. Data were band-pass filtered with cutoff frequencies of 0.15 and 28 Hz before downsampling to 64 Hz. Correction for eye blinks and ocular movements was performed by applying the standard subtraction algorithm based on correlations [Quilter *et al.*, 1977] to the average of the differences between channels Fp1–F1, and Fp2–F2.

3.2.1.1 Ocular artefact correction

The method introduced by Quilter *et al.* [1977] to remove artefacts from the EEG caused by eye movements is based on a least squares regression function that calculates the proportion of one variable (in this case, EEG) that is explained by another (ocular movements and eye blinks).

Formally, if E and X_i represent the EOG and the voltage in channel i , respectively, then B_i , the proportion of EOG in channel i can be estimated as the correlation between both channels, ρ , scaled by their standard deviations:

$$B_i = \rho \cdot \frac{SD(X_i)}{SD(E)}.$$

Correction of the EEG channels takes places as per

$$estEEG_i(t) = X_i(t) - B_i \cdot E(t) - C_i,$$

with C_i being used to remove the baseline effect from the EOG:

$$C_i = \overline{E(t)} - \overline{X_i(t)} \cdot B_i.$$

Since in this experiment we did not use EOG sensors, we estimated the EOG channel as the average of the differences between frontal channels Fp1–F1 and Fp2–F2.

3.2.2 Experimental Design

The images for the experiment consisted of 2,400 aerial pictures of London. All images were converted to grayscale and their histograms were equalised. Picture size was 640×640 px² (subtending 11.5×11.8 degrees of visual angle). Target (T) pictures were aerial pictures in which a randomly rotated and positioned airplane had been (photo-realistically) superimposed. Non-target (NT) images did not contain airplanes. Figure 3.1 shows examples of target and non-target images. The centroid of the targets was positioned at a maximum visual angle of 3.7 degrees in the horizontal axis, and 3.8 degrees in the vertical (with respect to the centre of the screen).

Pictures were shown to participants in sequences (or bursts) of 100 images which were presented at rates varying between 5–15 Hz, with no gaps between two consecutive stimuli as illustrated in Figure 3.2. Given the presentation rates, this is considered to be an RSVP protocol. Ten target pictures were randomly inserted within each sequence (the remaining 90 being non-targets) with the only restriction that there had to be at least one non-target image between two targets. Thus, the ratio of target vs non-target images was 10%.

Two sets of stimuli were generated: one that always contained the same airplane in target pictures and another in which the airplane for each target picture was randomly chosen (with replacement) from a set of three with different shapes

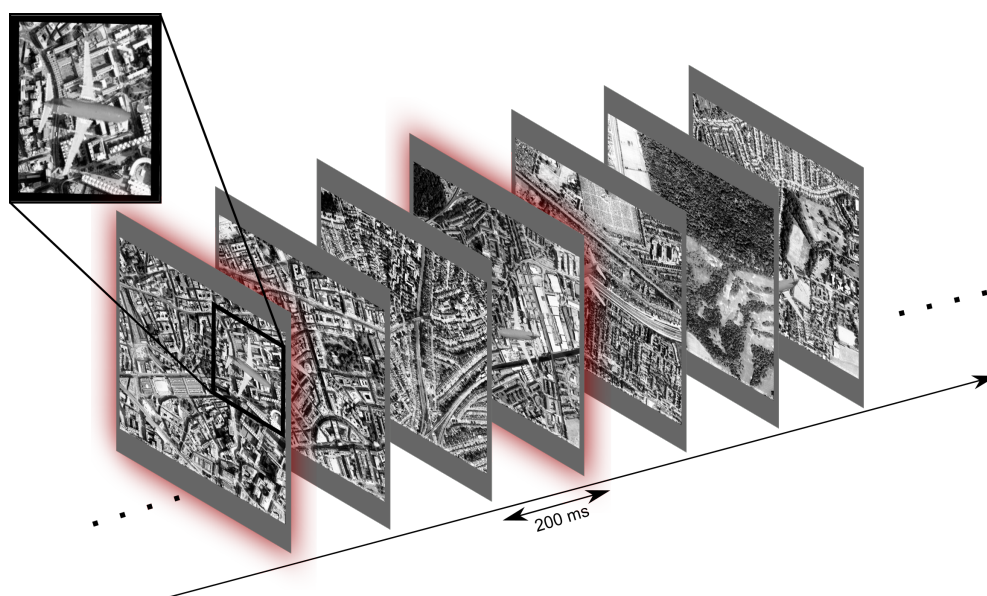
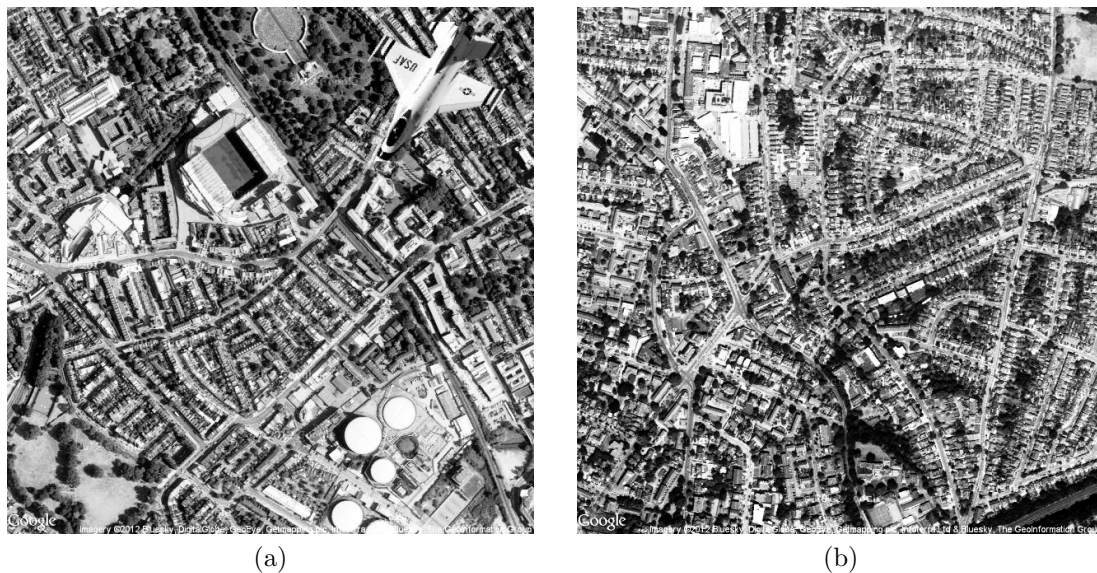


Figure 3.2: Illustration of the protocol used in the experiment, for a presentation rate of 5 Hz. For clarity, target images are highlighted in this figure.

and sizes. Figure 3.3 shows the three templates that were used on this experiment, extracted from some of the real images used in the experiment.

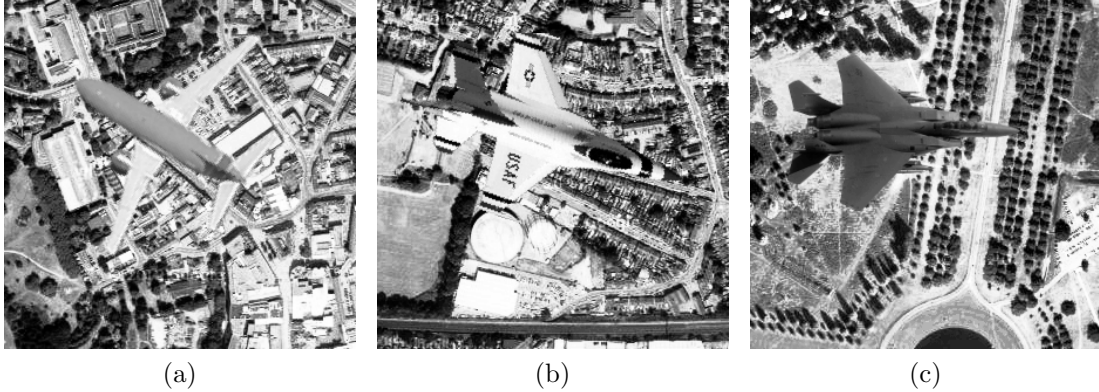


Figure 3.3: Target templates used in the RSVP paradigm. The template in (a) was used in all difficulty levels. Templates (b) and (c) only appeared in levels 3 and 5. Note that these are not the full images shown in the experiment, but fragments of them.

Table 3.1: Parameters of the different levels of the experiment.

	<i>Level 1</i>	<i>Level 2</i>	<i>Level 3</i>	<i>Level 4</i>	<i>Level 5</i>	<i>Level 6</i>	<i>Level 7</i>
Presentation rate (Hz)	5	6	6	10	10	12	15
# different targets	1	1	3	1	3	1	1
Burst duration (sec)	20	16.67	16.67	10	10	8.33	6.67

Participants were presented with bursts, at 7 levels of difficulty. The parameters for the levels of difficulty are given in Table 3.1. Each level consisted of 24 bursts of images, each composed by 100 images. In order to enhance the amplitude of the P300 [Polich, 2004a], participants were assigned the task of mentally counting the planes they saw within each burst and report the total at the end of a burst (to encourage them to stay focused on the task) [Yazdani *et al.*, 2010]. Participants could rest after bursts and were free to decide when to start the next burst (by clicking on a mouse button).

Despite the fact that we expected saccades to be greatly reduced at the presentation rates considered [Neider *et al.*, 2013; Potter & Levy, 1969], we instructed participants to try to reduce eye movements and blinks in order to obtain EEG

signals with as few artefacts as possible. All participants completed the experiment within 90 minutes.

3.2.3 Signal Processing and Feature Selection

Based on previous research (e.g., [Gerson *et al.*, 2006; Parra *et al.*, 2008]), the classification of target and non-target images was expected to rely mostly on the P300 ERP. For this reason, following the onset of each picture on the screen, epochs containing the 300–600 ms interval after stimulus onset were extracted. This resulted in a total of 20 samples (i.e., features) per electrode at the sampling rate of 64 Hz. Epochs were referenced to the mean voltage of the 200 ms interval before stimulus onset. The samples extracted from each electrode were concatenated to form feature vectors.

In an effort to reduce the total number of features used for the task and minimise the risk of overfitting, only centro-posterior-occipital electrode sites were used, as these are typically where P300s are most prominent. We used 28 electrodes (Oz, POz, Pz, CPz, CP1–CP6, TP7–TP8, P1–P10, PO7–PO8, PO3–PO4 and O1–O2 — see Figure 3.4), thus resulting in feature vectors of 560 elements for each trial.

3.2.4 Single-User Detection of Targets

For each level of difficulty, we used 10-fold stratified cross-validation to train an ensemble of two hard-margin linear Support Vector Machines (SVMs) for every participant. The stratified cross-validation method ensured that the proportion of targets and non-targets remained constant (at 10%) between training and testing

Trials were labelled as +1 for the epochs corresponding to a target image and -1 for non-target ones. The analogue outputs of the SVMs were used as scores: the higher the output, the more likely a picture contained a target. We used the Area Under the Receiver Operating Characteristic Curve (AUC) as a measure of the performance of the classifier (and, thus, of the BCI) [Bradley, 1997; Hanley & McNeil, 1982]. The value of the AUC of a perfect classifier is 1, whereas its value for a random classifier is 0.5¹.

In order to obtain the performance of the single-user BCI, for each participant and difficulty level the average AUC values across the 10 folds are reported. This method is labelled as sBCI (Single-user BCI) in the following figures and tables.

Due to technical difficulties with the recording equipment, the data for some participants and levels could not be recovered and used in the analysis, so whereas levels 1 and 3 contain data from all ($N=11$) participants, only data from 10 individuals are available for the remaining levels.

3.2.5 Collaborative Classification

Due to the fact that all the subjects had been exposed to the same pictures in the same sequence throughout the experiment, after the experiments it was possible to simulate the conditions of groups of participants performing the task simultaneously together with a collaborative BCI. Since 11 participants were available (see note at the end of the previous section), simulated groups were composed of up to 11 members.

¹Note that the accuracy is not an appropriate measure of classifier performance for imbalanced datasets: an accuracy of 90% in a dataset that contains a class ratio of 10:1 could mean that the classifier is always outputting the majority class [Kononenko & Bratko, 1991; Kubat & Matwin, 1997; Swets *et al.*, 1988]. The AUC, on the other hand, does not take into account the class probabilities and is independent of their frequencies [Swets *et al.*, 1988].

Several methods to merge the features of groups of users were studied:

- *Single-Classifier Collaborative BCI (SC-cBCI)*: this consisted of creating group epochs by averaging the feature vectors across the group members and training a single classifier using the average vectors as inputs.

In BCIs that operate based on the oddball paradigm, such as Donchin’s matrix speller [Farwell & Donchin, 1988], every stimulus is typically flashed several times before making a decision on what the user’s intention is. By averaging the epochs corresponding to these repetitions, a reduction of the noise level is achieved and the ERP appears clearer, thus achieving higher system accuracy. By adopting a single-classifier approach to collaborative BCI, a similar reduction of noise is obtained by averaging the signals (i.e., features) across multiple users instead of across multiple trials from the same user. However, given the inter-subject variability in the ERP waveforms, in principle, this method could also reduce the amplitude of the P300s and make the classification task more difficult.

- *Multiple Classifier Collaborative BCI (MC-cBCI)*: in this method, we averaged the outputs (i.e., scores) of individually-tailored classifiers (trained as described in Sec. 3.2.4).

Since the raw outputs of the classifiers are not binary, but proportional to the certainty that the system has of the presence of a target, the resulting MC-cBCI system can be seen as a voting algorithm that gives the same weight to all users.

- *Linear Discriminant Analysis-based MC-cBCI (LDA-cBCI)*: here, a Linear Discriminant Analysis (LDA) classifier was trained with the outputs of the

individually-tailored classifiers of the training set (see below). In this approach, the scores can be seen as votes from the group members, but the linear classifier may now assign different weights to different members if this improves group performance. LDA classifiers have the advantage that they do not have any parameters that need to be tuned, and given the very low number of features used in this case (which is equal to the number of group members), it seemed a sufficiently powerful candidate.

In all approaches, the 2-SVM ensemble received the same data as for the single-user case.

Additionally, in the case of the LDA-cBCI method, for each fold, the system was trained as follows: first, the training set was used to train the SVM classifiers individually for each user in a group. The concatenation of the outputs from the SVM ensembles for the training set across all users were used as the feature vector to train the LDA classifier. If this method (as opposed to using a separate set for training the LDA classifier) has any consequence on the final results, it would be overfitting on the training set of each fold, which would result in lower results on the test set, hence imposing a lower boundary on our results. However, due to the small amount of data available (specially of the target class) and after discarding oversampling as an option, we felt this was a sensible approach.

3.3 Results

This section presents the results of the aerial picture RSVP experiment. We begin by analysing the plane counts reported by the users at the end of each burst and then continue the study by presenting an ERP analysis based on the

Table 3.2: Average total plane counts reported by participants as a function of difficulty level. The total real count was 240 airplanes for all difficulty levels. For a quick reference, the presentation rate and number of different targets are also reported.

Level	1	2	3	4	5	6	7
Plane count	197.2	186.7	151.4	157.2	118.6	143.0	100.7
Sensitivity	82.2%	78.8%	63.1%	65.5%	49.4%	59.6%	41.2%
Presentation rate (Hz)	5	6	6	10	10	12	15
# different targets	1	1	3	1	3	1	1

epochs extracted from the EEG recordings. This section finishes with the results of the single and collaborative BCIs.

3.3.1 Behavioural Results

As was described in Section 3.2.2, participants were asked to provide the number of targets they had seen at the end of each burst. Even though these numbers were not taken into account for the machine learning part of the experiment, it is instructive to look at them to ascertain the objective difficulty of each level.

Evidence indicating that the difficulty of the task increases with the rate of presentation and the number of target templates used in a level is provided by these records of reported plane counts. Indeed, as shown in Table 3.2, average plane counts decrease as the presentation rate increases. At 12 Hz, more than 40% of targets were missed by participants.

Moreover, levels with three types of targets (i.e., levels 3 and 5) are also more difficult for the users than their respective single-target counterparts (levels 2 and 4, respectively), which present sensitivities about 15% higher for the same presentation rate.

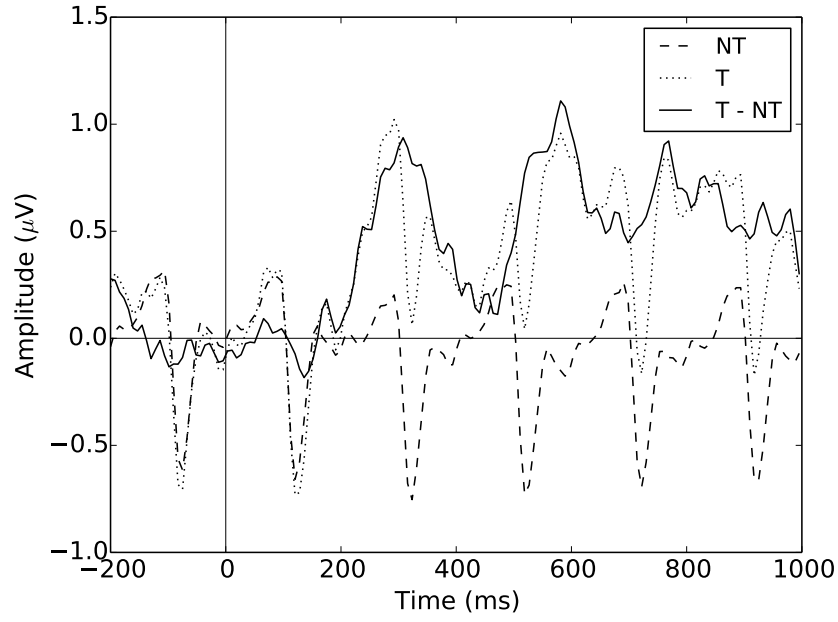
3.3.2 ERP Analysis

Before we look at the ERPs obtained in the experiments, it should be noted that, as is common in other BCIs such as the matrix speller [Cinél *et al.*, 2004; Farwell & Donchin, 1988], when using very short SOAs, ERPs are significantly deformed with respect to their “textbook” form found in electrophysiology and neuropsychology studies [Luck, 2005; Woldorff, 1993]. For instance, in this RSVP experiment, EEG signals contain a large SSVEP component at the frequency of stimulation due to the involuntary response of the visual system. However, as we will see below, this waveform is modulated by the ERPs selectively generated by different stimuli.

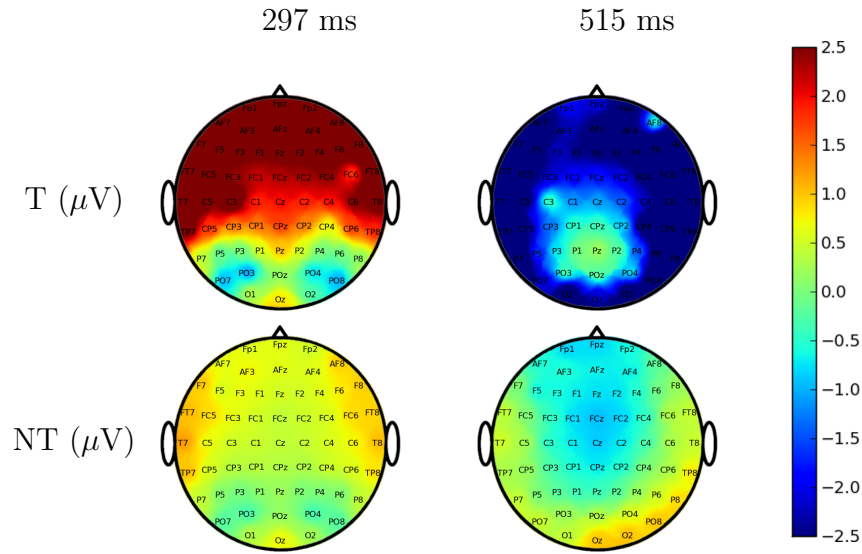
Let us start by looking at the grand averages of the ERPs for the T and NT conditions, for a presentation rate of 5 Hz. These, together with their difference, are shown in Figure 3.5 (a) for electrode site Pz, while Figure 3.5 (b) shows two snapshots of the corresponding scalp distributions, taken at approximately 300 ms and 515 ms after stimulus onset.

There appear to be two peaks on the difference waveform of Figure 3.5 (a), one at about 300 ms and a second one around 600 ms, both referred to stimulus onset. However, the P3b, which we expected to see in response to the stimuli, should not be double-peaked [McCarthy & Donchin, 1981; Polich, 2004a, 2007]. The valley between the two peaks could be due to this SSVEP modulation, the effects of which are also manifested in the scalp maps reported on Figure 3.5 (b).

To see if the observed differences between target and non-target evoked responses were statistically significant, the peak amplitude of the “first” peak was defined as the mean voltage amplitude in the time intervals 300–400 ms after stim-



(a)



(b)

Figure 3.5: (a) Stimulus-locked grand averages for T and NT trials at channel Pz and their differences and (b) scalp maps of the grand averages for T and NT at 297 ms and 515 ms after stimulus onset, for a presentation rate of 5 Hz.

ulus onset, and the peak amplitude of the second peak (P3b) as the mean voltage amplitude 500–600 ms after stimulus onset for T and NT trials. A Kruskal-Wallis test (a one-way, non-parametric, analysis-of-variance type of test) was applied to test for differences between the T and NT conditions for electrode sites Cz, CPz and Pz for both peaks. All such differences were found to be highly significant at the tested electrodes (p values $< 3 \times 10^{-4}$ after Bonferroni correction) except for the P3b peak-amplitudes in Cz, which are not statistically different.

3.3.3 Single-User BCI

Let us begin with the performance of the system when the signals from a single user are used for the classification of images into targets and non-targets. As was mentioned previously, the AUC was used to measure the performance of the machine learning component.

In order to check what would the performance (measured by the AUC) for a random classifier be, we ran a test in which we shuffled the labels of the data and then performed the cross-validation step as described above. Averaging across all participants and levels of difficulty, we obtained a mean $\text{AUC} \pm \text{SD} = 0.5 \pm 0.02$. This was expected, given that, unlike the accuracy, the AUC is insensitive to class imbalance.

Moving now to the actual results, Figure 3.6 reports the boxplots of the AUCs for every level of difficulty, across all volunteers. These are consistent with previous published results: user performance decreases with presentation rate and the best performance is associated with the first (easiest) difficulty levels [Yazdani *et al.*, 2010]. Another expected result is the big drop in AUC when moving from

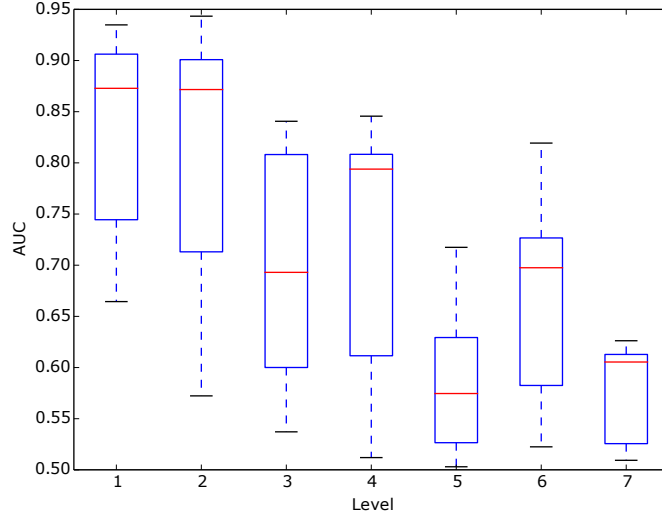


Figure 3.6: Distributions of mean AUCs across the cross-validation folds for single-user BCI across all levels of the experiment.

only one type of plane in target images to three types of planes, possibly due to some of the new airplane templates being more difficult to discern, or due to habituation effects to the main plane template presented in most levels. These results are in agreement with the corresponding decrease in reported plane counts when going from level 2 to level 3 (both at 6 Hz) and from level 4 to level 5 (both at 10 Hz).

Less expected was the fact that even at the very high rates of presentation used in the last two levels of difficulty (12 and 15 Hz, respectively), the median AUC value across all subjects (represented by red horizontal lines in the boxplots) is nowhere near 0.5 (the performance of a random classifier measured by the AUC), indicating that the participants' visual system was still able to discriminate, at least to some degree, targets from non-targets, despite the very low plane counts from Table 3.2, which indicate that the labels for the data used to train the classifiers may have been heavily contaminated with missed targets.

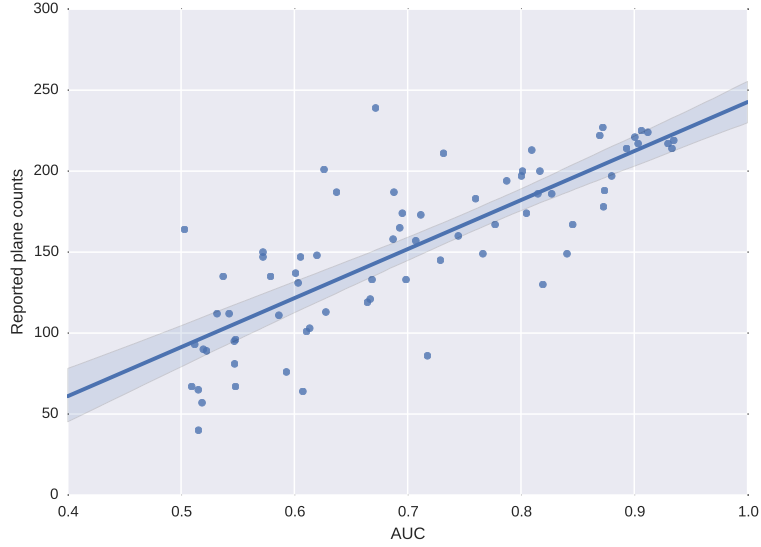


Figure 3.7: Scatterplot of the reported plane counts for each participant and difficulty level vs the corresponding AUC obtained in cross-validation for that participant and level of difficulty.

In order to get a measure of how well the classifiers perform with respect to the reported plane counts, we calculated the correlation coefficient between the mean cross-correlation AUC from each participant and for each level with the reported plane count from that participant and level. This resulted in $\rho=0.8$ (p value = 3.9×10^{-17}), showing that as expected, performance increases when the person reported seeing a number of planes closer to the real one. Figure 3.7 shows a scatterplot of these two variables, together with the regression line ($R^2=0.64$). Shaded areas represent the 95% confidence interval for the regression line, obtained by bootstrapping the data 1000 times.

The boxplots from Figure 3.6 show a wide range of AUCs within each level. Also, the distribution of AUCs across the population of volunteers is clearly skewed for most levels, as evidenced by the non-symmetrical boxes around the median value. A few participants performed much worse than the average of the

group. In particular, participant 4 was over-zealous: at the end of the experiment he reported that he had blinked to signal each target (even though he had explicitly been told to try to avoid blinks!). Participant 5, instead, was drowsy: he regularly reported a number of targets seen in each burst much smaller than the real value. Also, when we plotted the averages of the signals for targets and non-targets for this participant, we found that there were almost no differences between them.

In normal conditions, BCI researchers would exclude such participants as they did not behave as was required by the experiment. However, this will not be done here, as it actually serves the purpose of illustrating the benefits of a collaborative BCI.

3.3.4 Collaborative BCI

As mentioned before, several methods of aggregation were tested in order to create a collaborative BCI system: the single-classifier cBCI (SC-cBCI), multiple-classifier cBCI (MC-cBCI) and LDA-based MC-cBCI (LDA-cBCI) approaches. Figures 3.8–3.10 show the average AUCs across all possible combinations of groups and group sizes for each level and method.

The first feature that can be observed from these figures, and which is common to all of them, is that bigger groups lead to higher AUCs. However, the gains obtained from adding extra members to a group progressively reduce. This effect is also observed, for example, in matrix spellers, where increasing the number of repetitions after a certain number does not improve accuracy — on the contrary, it might slow down the BCI unnecessarily and make users get tired sooner.

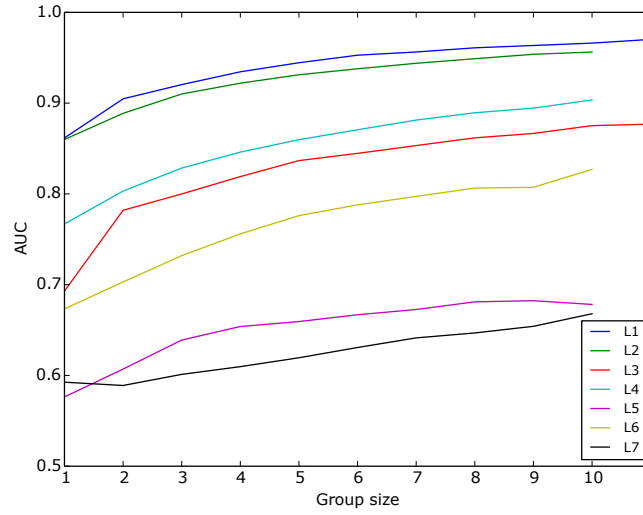


Figure 3.8: Median AUC values across participants (for size 1) and all possible combinations of participants plotted for every level when using the SC-cBCI method.

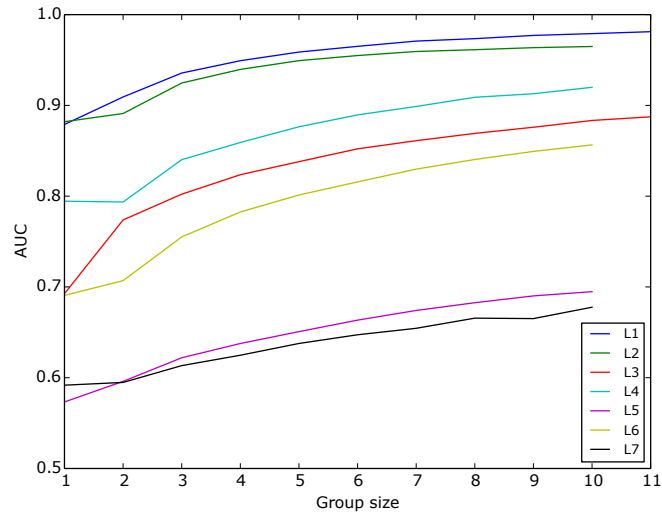


Figure 3.9: Median AUC values across participants (for size 1) and all possible combinations of participants plotted for every level when using the MC-cBCI method.

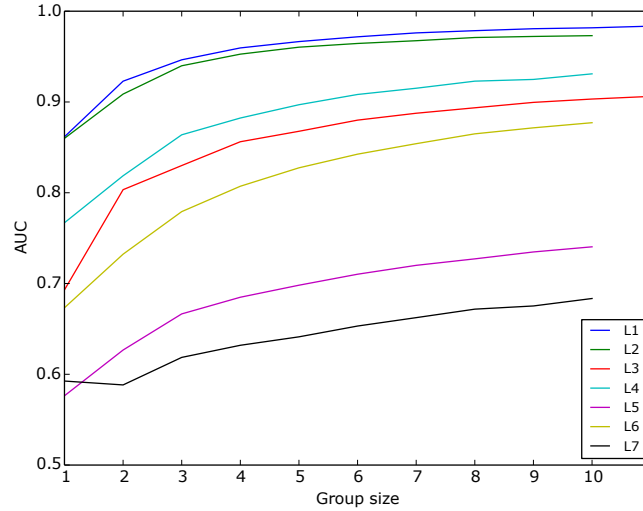


Figure 3.10: Median AUC values across participants (for size 1) and all possible combinations of participants plotted for every level when using the LDA-cBCI method.

In the case of cBCIs, this seems to happen for groups bigger than 6–7 members for the easier levels, although for more difficult ones (i.e., levels 6 and 7) the performance of the cBCI can still be improved, on average, by increasing the size of the groups. However, for reasonable presentation rates (e.g., level 4, at 10 Hz), the cBCI can still get very close to a perfect classifier, especially for the LDA-cBCI method.

In order to compare the different methods, for each level and group size one-sided pairwise comparisons were performed between SC-cBCI, MC-cBCI and LDA-cBCI methods by means of a paired Wilcoxon test. The full results and p values from this test can be found on Appendix B. Table 3.3 summarises these results (after applying Bonferroni correction for multiple comparisons). Out of the 195 comparisons that could be performed, 139 were statistically significant (after applying the correction).

For a group size of 2 at level 5, the LDA-cBCI method was significantly better

Table 3.3: Summary of comparisons between the 3 modes for combining evidence to create collaborative BCIs: LDA-cBCI (“L”), MC-cBCI (“M”) and SC-cBCI (“S”). The symbol $>$ is used to indicate that the method on the left is statistically significantly better than that of the right. The symbol $=$ indicates that the methods on both sides are not significantly different. N/A values represent those cases in which there were not enough samples to perform the test.

<i>Group size</i>	<i>Level 1</i>	<i>Level 2</i>	<i>Level 3</i>	<i>Level 4</i>	<i>Level 5</i>	<i>Level 6</i>	<i>Level 7</i>
2	L>M=S	L>M=S	L>M=S	L>M=S	L>M	L>M=S	L=M=S
3	L>M>S	L>M>S	L>M>S	L>M>S	L>S>M	L>M>S	L=M>S
4	L>M>S	L>M>S	L>M>S	L>M>S	L>S>M	L>M>S	L>M>S
5	L>M>S	L>M>S	L>M>S	L>M>S	L>S>M	L>M>S	L>M>S
6	L>M>S	L>M>S	L>M>S	L>M>S	L>M=S	L>M>S	L>M>S
7	L>M>S	L>M>S	L>M>S	L>M>S	L>M=S	L>M>S	L>M>S
8	L>M>S	L>M>S	L>M>S	L>M>S	L>M=S	L>M>S	L>M>S
9	L>M>S	N/A	L>M>S	N/A	N/A	N/A	N/A

than the SC-cBCI method. However, MC-cBCI was not found to be significantly different from any of them.

It is obvious from Table 3.3 that the LDA-cBCI method is statistically significantly better than the other two, which was to be expected since it is trained to assign optimal weights to the vote of each group member. From the results shown in this table, it is also very easy to rank the three methods: LDA-cBCI is overall significantly better than MC-cBCI, which is, in turn, significantly better than SC-cBCI. The fact that MC-cBCI performs better than SC-cBCI is not surprising either, since different participants may have different P300 latencies and their ERPs could have different shapes, so averaging the preprocessed EEG signals (as is done in the SC-cBCI method) does not always contribute to decreasing the signal-to-noise ratio (SNR) in a way that is beneficial for the BCI.

3.4 Discussion

This chapter presented the experiment that will be used through much of this PhD thesis and studied the possibility of using a BCI for the automatic classification of images that contain objects or features of interest presented at a very high rate. It also examined different ways of combining the observations across several users in order to improve the performance of the single-user BCI.

It was mentioned at the beginning of this chapter that the only objective measure of how good participants were at detecting the targets was the plane counts that they reported at the end of each burst. There were very few occasions where these counts were in agreement with the actual number of planes that had been inserted in the burst. Thus, one of the difficulties faced when training the classifiers was not having a “clean” set of data as is common in BCI research. Indeed, those levels for which the participants reported lower plane counts also showed poor BCI performance (although still above chance levels), and we found that the reported counts were highly correlated with the performance of our classifiers for single user BCIs.

We showed that levels 3 and 5, for which three different target templates were given to the participants, had much lower plane counts than levels 2 and 4, respectively. This could be due to either participant habituation to the main target template, or to the new templates being harder to discern from the aerial image background. Visual search is known to be affected by target eccentricity [Gruber *et al.*, 2014] and this, in turn, affects neural classification [Dias & Parra, 2011; Marathe *et al.*, 2016]. However, the distribution of target positions around the x and y axes was not different between the 1-plane and the 3-plane levels.

Another difficulty present in this work is the environment in which the signals were collected. Whereas BCI research is usually conducted on data recorded in rooms that are shielded from electromagnetic noise, these experiments were carried out in a normal room (with no shielding) where the volunteer could be subject to disruptions and general noise at any point. Still, considering these and the previously highlighted factors, the system achieved an acceptable classification rate (for a presentation rate of 10 images/second, a median AUC around 0.8 was obtained in level 4) for single-user BCIs.

In the collaborative case, we merged signals from groups of users using several different methods. In all cases, the performance of the classifiers was better than in the individual case, and increasing group size improved the average group performance. Studies on decision making have shown that bigger groups lead to better decisions. In this case, we simulated groups of up to 11 members, and showed that cumulative improvements decrease with increasing group sizes. This might be due to the fact that bigger groups are less affected by “noisy” members. In a way, the information added by the extra members acts as the trial-averaging technique in normal BCIs for disabled users. Thus, as was the case for the matrix spellers, the SNR is no longer improved by extra averaging (i.e., adding new members to the group).

One possible limitation of the work presented here is that the results were obtained through a cross-validation loop, and no proper test set was used in this chapter (a practice that is becoming increasingly popular in BCI research). Hence, although it is unlikely that our results are due to overfitting to the dataset, the real performance of the system needs to be assessed in the future using unseen data.

With respect to the results obtained with the LDA-cBCI approach, given the fact that the LDA was trained using the SVM response to trials from the *training* set from each fold, we believe that there is a chance that the LDA classifier that was used to assign different weights to the different users may have overfitted during training. If this was the case, then the actual performance when trained with different data, perhaps from an additional set of trials that are not part of the training set for the SVMs, would likely be higher than the one reported here. We did not attempt this due to the limited amount of data of the target class available. However, it is another matter that should be explored in the future.

In the collaborative BCIs, we observed the same influence of number of target templates on performance as in the single-user BCIs (i.e., levels with more than one type of target showed marked decreases in performance with respect to the equivalent levels that only contained one airplane). This effect is heavily correlated with the reduced plane counts reported by the participants, so the machine learning component of the BCI has very few target cases in which the participant did indeed see the target in the picture, which adds noise to the already heavily imbalanced training data. However, for the presentation rate of 6 Hz (level 3), it is still possible to reach good performance by means of a collaborative BCI. Since the RSVP protocol has been previously used to identify targets that were not fixed, we believe that the reason behind this decreased performance has to do with participants getting used to the 1-plane paradigm, and not to the multiple target scenario. However, this will have to be further studied in the future.

It should be reiterated here that there were two participants that dragged the performance of the single-user BCI down. Many of the groups that were formed contained data from these participants and so also cBCI performance was

affected, albeit to a lesser degree.

3.5 Conclusions

There is certainly an interest in the use of the RSVP technique to classify images without the user having to manually select the interesting ones. Here, an RSVP–BCI paradigm was applied to the field of target detection in broad area search, both in single-user and collaborative BCIs. However, there are other fields where this method could be useful. For instance, there is a vast amount of medical images that have to be seen by skilled clinicians on a daily basis for clinical purposes. Hope *et al.* [2013] applied the concept to the screening of mammographies by experts. In medical imaging, a collaborative BCI would be of great help, since a diagnosis can be subjective to a professional. By averaging across multiple experts, part of this bias could potentially be eliminated.

Last, but not least, finding the pictures that contain events of interest is only the first step to having a complete system for target detection. The following chapters will build upon the system and the techniques presented here by using the N2pc ERP to locate the targets within pictures (Chapter 4), concatenating the target detection and the target location systems (Chapter 5) and studying methods for participant selection when forming groups for collaborative BCIs (Chapter 6).

Chapter 4

Single and Collaborative BCI for Target Localisation

The N2pc event-related potential appears on the opposite side of the scalp with respect to the visual hemisphere where an object of interest is located. This chapter explores the feasibility of using it to extract information on the spatial location of targets in aerial images, both in single and collaborative BCIs. As a byproduct, the chapter also reveals an interesting relationship between handedness and the shape of the N2pc ERPs that are evoked by lateralised targets.

4.1 Introduction

The previous chapter studied the feasibility of detecting targets in rapid streams of images. Detecting targets accurately and at high speeds, however, is often only a prerequisite to more sophisticated processing. For instance, triage systems would benefit from techniques, such as the one that will be presented in this

chapter, that could automatically establish the *position* of targets within the images. While the P300 ERP is one of the most widely used ERPs for controlling BCIs (both in the traditional and the newer paradigms) and recognising targets, it does not allow this more sophisticated task.

The N2pc ERP is usually reported in literature to be related to selective attention processes. This chapter explores the feasibility of using this component to extract information on the spatial location of targets. Moreover, we will do this using stimuli representing complex real-life scenes for a task of practical utility (aerial image sifting) presented using the RSVP protocol described in the previous chapter through BCIs based on single-trial classification. This is a very difficult task for a BCI as: (1) the N2pc is a much smaller ERP than the P300 in terms of voltage amplitudes, duration and locations where it can be detected, (2) we use a single-trial approach, and (3) images are presented at a rapid presentation rate.

The most similar work to the one we explore here is that of Putze *et al.* [2013], where the authors used EEG data to detect the targets and eye tracking for locating them by asking participants to fixate their eyes on them. However, this was not done on an RSVP task, but rather on a series of simple stimuli (a number of circles arranged in a circle) that were sequentially and randomly flashed. Moreover, it is not clear whether participants would be able to fixate on the target at the high speeds used in RSVP (Putze *et al.* [2013] flashed each stimulus for 2 s), as there are previous reports of saccades being suppressed at such rates [Neider *et al.*, 2013; Potter & Levy, 1969].

Thus, if the N2pc can be detected in a scenario such as the one presented in the previous chapter, it could be exploited, for example, to help circumscribe the

area of the image where the target is located, thereby speeding up the job of the person reviewing the potential targets detected by a BCI without the need for the extra equipment and calibration of the eye tracker. Also, it could help improve target detection if, for instance, targets too lateral with respect to an observer's gaze to elicit a fully blown P300 [Dias & Parra, 2011; Marathe *et al.*, 2016] still cause a detectable shift of attention resulting in an N2pc. To verify these hypotheses, we also investigated the relative dependency of target classification on the P300 and the N2pc. Moreover, in Chapter 5 we will combine the target detection system presented in the previous chapter and the target localisation system presented here to see whether the P300 and N2pc can help each other in detecting and locating targets.

This chapter is organised as follows. Section 4.2 describes the experimental setup, the signal acquisition and manipulation and the methods used for feature selection and classification. Section 4.3 reports on (and discusses) the results of the experiments, both in terms of ERPs and of the localisation accuracy of the BCI. This section also reports the handedness results obtained as a byproduct of these experiments. We discuss the implications of these findings in Section 4.4 and provide some conclusions and indications for future work in Section 4.5.

4.2 Methods

4.2.1 Participants and Setup

The work presented in this chapter uses the signals acquired from the pool of participants described in Chapter 3 (see Section 3.2.2). The signals were also

Table 4.1: Distribution of lateral targets across levels. The reported percentage of lateral targets is referred to the total number of targets within the level — there are a total of 240 targets per level.

Level	Lateral Targets	LVF	RVF
1, 2, 4, 6, 7	60%	59	85
3, 5	70%	90	79

preprocessed in the form described in that chapter (i.e., band-pass filtered between 0.15–28 Hz and downsampled to 64 Hz). However, this chapter focuses on different electrodes and time windows, as will be described below.

4.2.2 Experimental Design

A full description of the experiment and participants can be found in Section 3.2.2. In this chapter we will focus on a specific subset of the data: those trials that correspond to target pictures and, more specifically, those that contain a lateral target. Concatenating the target detection and target localisation systems and studying the influence of non-targets on the localisation system will be done in the next chapter.

The horizontal position of a target was defined as the x-coordinate of the centroid of the plane contained in the image. Lateral targets were those the centroid of which was positioned at least at a visual angle of ± 1.2 degrees on the horizontal axis (with respect to the centre of the screen).

The numbers of lateral targets and their distribution between Left Visual Field (LVF) and Right Visual Field (RVF) targets for each level are given in Table 4.1¹.

¹The imbalance in the cardinality of the LVF and RVF target sets is due to a slight undetected bias in the algorithm that was used to position the planes. This was, however, inconsequential other than it slightly reduced the statistical significance of some of our findings.

4.2.3 Feature Selection and Classification

The following subsections will describe different uses of the experimental data, each characterised by a different choice of training and test sets and methods. In particular, as highlighted in previous sections, we will start by verifying the presence of the N2pc component in the conditions of the RSVP experiment (Section 4.2.3.1), and using this ERP to determine whether a lateral target appeared on the left or the right side of the screen. We will then study differences in the N2pc due to the preferred hand of the participants (Section 4.2.3.2), and use these to improve the target localisation task by being able to predict the actual coordinate of the airplane (Section 4.2.3.3). Following the structure from the previous chapter, we will then study the three approaches of combining evidence from multiple users to create cBCIs (SC-, MC- and LDA-cBCI) in the form described in Section 4.2.3.4. Finally, the influence of the N2pc component on the target detection task is studied (Section 4.2.3.5).

4.2.3.1 Detection of the N2pc Component

To verify whether the N2pc component could be detected in single trials in the conditions of this experiment, epochs of EEG signal were extracted from approximately 200 ms to 400 ms after stimulus onset (the temporal window where the N2pc most often occurs according to the literature). This resulted in 14 samples per channel at the 64 Hz sampling rate used. The data were referenced to the mean value of the 200 ms interval before stimulus onset.

Since the N2pc is a lateralised ERP, it is more easily revealed when looking at *differences* between pairs of electrodes corresponding to symmetric positions with

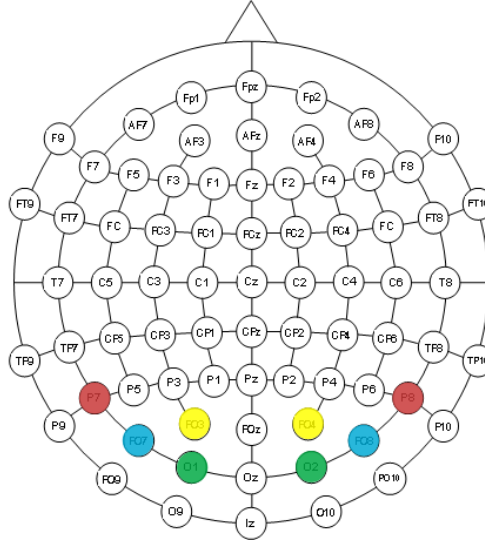


Figure 4.1: Electrodes used for target localisation (based on the N2pc).

respect to the brain’s median plane than when processing left and right electrodes independently. Furthermore, the N2pc is most prominent in the posterior and occipital electrodes. Based on this, when detecting N2pc components in this experiment we used the set of four differences between electrode pairs: (P7–P8), (P7–P8), (PO3–PO4) and (O1–O2) (highlighted in Figure 4.1, where each pair is represented by a different colour). Concatenating these electrode differences yields a feature-vector representation of epochs including $14 \times 4 = 56$ elements.

With this input representation, linear SVM classifiers were trained to distinguish between LVF and RVF targets following the same stratified 10-fold cross-validation approach described in Chapter 3.

For each level of difficulty, the analogue output scores of the SVMs were recorded, and used to compute the ROC curve for each participant. As before, the information contained in each ROC curve was condensed into the AUC score.

4.2.3.2 Handedness Detection

The pool of volunteers for this experiment had an approximate balance between left-handed (LH) and right-handed (RH) participants (5 out of the 11 participants were left-handed)¹. Linear SVM classifiers were trained to perform discrimination of participants based on handedness through an 11-fold leave-one-participant-out cross-validation loop. Each fold was trained with all the RVF target epochs from difficulty level 1 for all LH and RH participants except for the excluded one. After training, these trials were fed again to the classifiers, and the median of the raw output score was computed and used as a threshold for classifying the left-out participant as LH or RH.

The test set for each fold was composed by all the RVF target epochs from the excluded participant. After obtaining a score for each trial, the median score value from the test set was compared with the classification threshold obtained from the training set.

As we will see in Section 4.3.3, there are also differences in the N2pc component evoked by LVF targets which are associated with the handedness of the participant. However, only the RVF epochs were used because these differences are greater for right visual field targets than they are for targets located on the left.

¹By “handedness” we refer to the self-reported handedness of the participants in the study. More specifically, volunteers were asked for their preferred hand for writing. Since it was not expected that there would be implications of the study on handedness research, the standard tests routinely used to more objectively verify the handedness of participants were not used.

4.2.3.3 Target Localisation

After having determined that the N2pc can help separate left from right visual field targets, we explored the possibility of using it to tell to what degree a target is lateral with respect to the centre of the image. For this, a linear predictor which was optimised by a Particle Swarm Optimiser (PSO) [Poli *et al.*, 2007] was used. The representation used in the PSO included 17 parameters: eight of these were interpreted as indices in the 56-dimensional feature vectors extracted from each epoch (see Section 4.2.3.1), other eight were the coefficients for the corresponding features, and the remaining one was a constant term for the linear predictor. The fitness function optimised by the PSO was multi-objective as it aimed at: (1) obtaining a correlation, $\rho_{\text{predictor}}$, between actual outputs and desired outputs (the x coordinate of the target, in pixels, in the picture corresponding to each epoch) as close as possible to the correlation, $\rho_{\text{reference}}$, obtained by a standard linear regressor using all the features, and (2) ensuring the regression line between the desired outputs and the outputs of the linear predictor has as a slope close as possible to 1¹. Formally, the fitness function (to be minimised) was:

$$f = |1 - \text{slope}|^2 + |\rho_{\text{reference}} - \rho_{\text{predictor}}| + 0.0005 \times \text{MAE}$$

where **slope** and **MAE** are the slope and the mean absolute error of the linear regression between the desired outputs and the outputs of the linear predictor, respectively.

The training set for this task was composed by a random sample that con-

¹This approach was used to compensate for the tendency of standard multivariate regression to compress its output range in the presence of strong noise on its inputs.

tained 65% of all targets (and not only the lateral ones) from each participant. The test set was composed by the remaining target epochs. This task was performed separately for the groups of left-handed and right-handed participants.

4.2.3.4 Collaborative Classification

The three approaches for creating cBCIs described in Section 3.2.5 (SC-cBCI, MC-cBCI and LDA-cBCI) were used again in this chapter to merge signals from multiple participants. Given the fact that the N2pc is not as variable across participants as the P300, we expected that the SC-cBCI approach would be worthy of exploration in this context, despite the fact that it was always the worst of the three methods of collaborative BCI for target detection.

4.2.3.5 Influence of the N2pc on Target Detection

Based on the results reported in the previous chapter and the literature review, we expected the classification of target and non-target images to rely mostly on the P300 ERP, but we wanted to explore the degree to which the N2pc could influence it. Epochs containing the 300–600 ms interval (which is where typically P300s appear) after stimulus onset were extracted. At the final sampling rate of 64 Hz, this resulted in a total of 20 features per electrode.

As before, in an effort to reduce the total number of features used for the task and minimise the risk of overfitting, only centro-posterior-occipital electrode sites were used, as these are typically where P300s are most prominent. One combination (E_{28}) used the 28 electrodes that were used in Chapter 3 (i.e., all the electrodes highlighted in Figure 4.2 — Oz, POz, Pz, CPz, CP1–CP6, TP7–TP8,

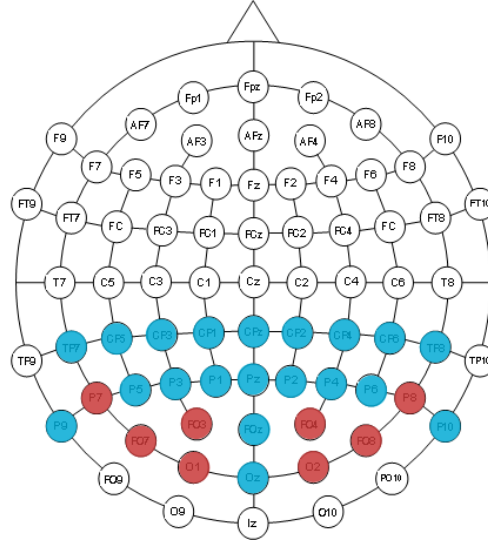


Figure 4.2: Electrodes used for evaluating the influence of the N2pc on target detection.

P1–P10, P7–P8, PO3–PO4, O1–O2). The second combination (of 20 electrodes, E_{20}) was identical to the first except that it omitted the electrode sites that were used for N2pc detection (i.e., only the electrodes shown in blue in Figure 4.2 were used). A third combination (E_{24}) included the electrodes in E_{20} plus the four electrode differences used for target localisation (i.e., the electrodes pictured in blue in Figure 4.2, plus the four differences highlighted in Figure 4.1). So, for the purpose of classification, epochs were represented with between 400 and 560 features.

The training and test sets used in the stratified 10-fold cross-validation loop included all the trials for each level (since the task here was target detection). As before, epochs were referenced to the average voltage in the 200 ms interval before stimulus onset.

4.3 Results

This section starts with an ERP study of the N2pc in order to show that this component is elicited in the conditions of this experiment. Then, we will report on the performance of single-user BCIs for single-trial LVF vs RVF classification. The section will continue by looking at the differences in the perception of lateral targets associated with the handedness of participants. Based on these differences, we study the degree to which a linear predictor can be used to quantify the eccentricity of a target with respect to the centre of the image. Finally, collaborative classification is performed by combining data from groups of users by means of the MC-cBCI, SC-cBCI and LDA-cBCI methods that were presented in Chapter 3.

4.3.1 ERP Analysis

The presence of the N2pc ERP in this experiment is illustrated in Figure 4.3, which shows grand averages for lateral targets for a presentation rate of 5 Hz. The “contralateral” line in the figure represents the grand average of participant averages that were computed with the epochs recorded from channel PO7 (on the *left* hemisphere of the scalp) for RVF targets with the epochs recorded from channel PO8 (on the *right* region of the scalp) for LVF targets. Similarly, the “ipsilateral” line represents grand averages computed from the epochs recorded at channel PO7 for LVF targets and the epochs recorded at channel PO8 for RVF targets. The adoption of these ipsilateral and contralateral grand averages follows the conventions of the N2pc literature (see Section 2.2.2), as these emphasise left-right asymmetries that would otherwise be lost with standard averages. Following

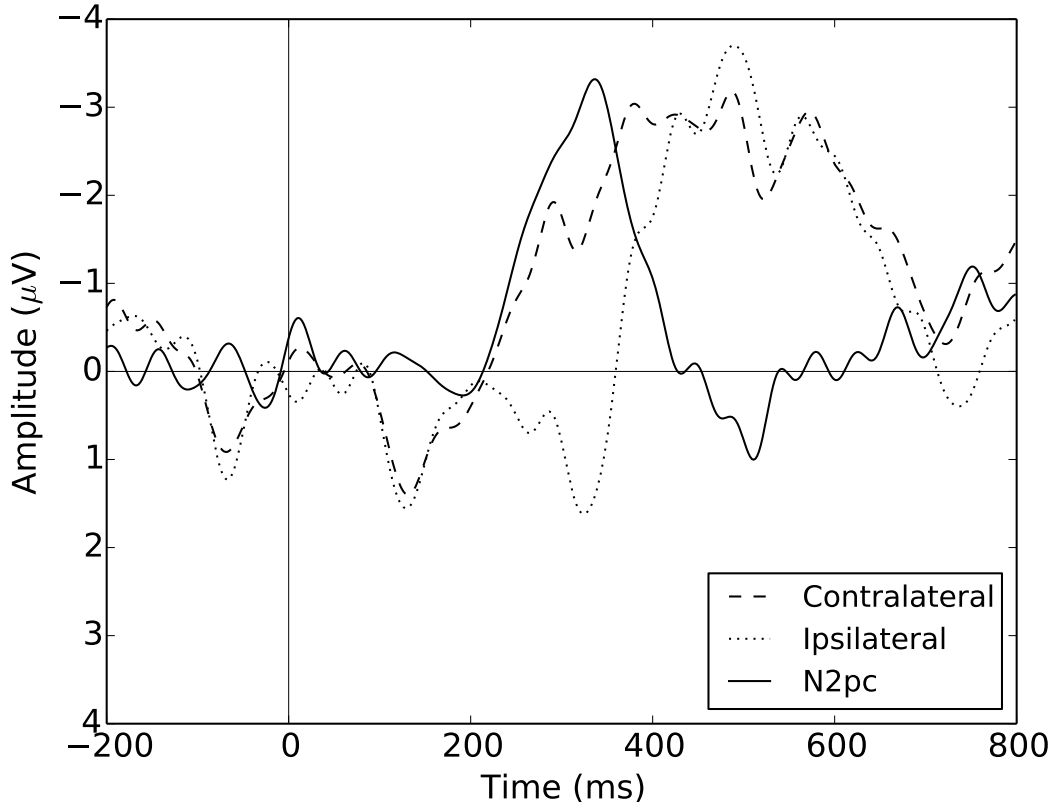


Figure 4.3: Contralateral and ipsilateral stimulus-locked grand averages at channels PO7 and PO8 and their difference (continuous line, “N2pc”) across lateral targets from the training set from one of the cross-validation folds.

the same conventions, the data were plotted using an *inverted ordinate axis* (so higher means more negative). To further illustrate the differences between the two conditions, the figure also reports the difference between the contralateral and ipsilateral grand averages (line labelled as “N2pc”). It should be noted that the SSVEP effect that could be observed in the P300 ERP plots presented in Figure 3.5 as well as in the contralateral and ipsilateral lines in Figure 4.3 is removed by performing this subtraction.

As we can see from the figure, the ipsilateral and contralateral ERPs start to deviate markedly from each other at 250 ms after stimulus onset, with their

difference peaking at approximately 340 ms. The shape and sign of the deflection is consistent with those of the N2pc reported in the literature [Eimer, 1996; Luck, 2012; Luck & Hillyard, 1994b], even though in this experiment its latency was slightly longer than in other studies, presumably because attention (both covert and overt) is also attracted (and, thus, divided) by features of the constant stream of distractors (non-targets) used in these experiments.

Moreover, the SPCN waveform (see page 16) that follows the N2pc in visual search that requires keeping the target in memory is not present in our experiment [Dell’Acqua *et al.*, 2006; Jolicœur *et al.*, 2006a, 2008; Thiery *et al.*, 2016]. This could be because the task we gave to the participants was just to count the number of targets, but we did not require them to memorise any aspect of them.

Figure 4.4 shows snapshots of the temporal evolution of the grand averages across the scalp between approximately 310 ms and 375 ms after the presentation of images containing a lateral target for the same low presentation rate of 5 Hz. Looking at the grand averages for LVF targets (top), the voltages at several contralateral posterior and occipital electrodes (e.g., PO4, PO8, P8) start becoming more negative than those in corresponding ipsilateral channels (e.g., PO3, PO7, P7) from around 300 ms after stimulus onset. This difference increases over time. The same effect can be observed in the grand averages for RVF targets (bottom), where the voltages at left posterior electrodes (i.e., the contralateral channels) are more negative than the corresponding voltages of the right (i.e., ipsilateral) channels in the same time interval.¹

The N2pc peak amplitudes for LVF and RVF targets were measured as the

¹Of course, there are asymmetries of brain function in the left and right hemispheres and, so, we cannot expect perfectly symmetric scalp maps.

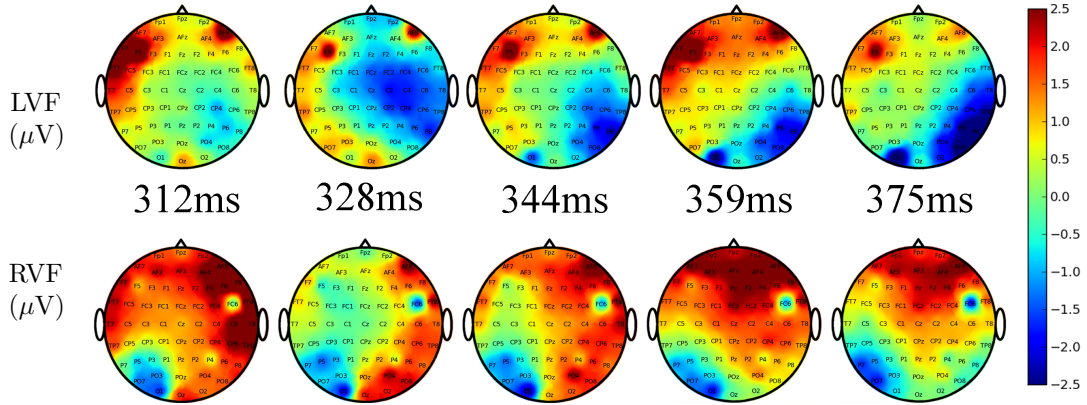


Figure 4.4: Grand-averaged scalp distributions between 312 ms and 375 ms after the onset of LVF (top row) and RVF targets (bottom row) at a presentation rate of 5 Hz.

Table 4.2: Medians and Kruskal-Wallis p values for the peak amplitudes of the voltage differences between contralateral and ipsilateral channels for LVF and RVF targets at a presentation rate of 5 Hz.

Electrode difference	LVF	RVF	p value
$PO7 - PO8$	$1.646 \mu V$	$-1.950 \mu V$	2.2×10^{-16}
$P7 - P8$	$1.897 \mu V$	$-1.301 \mu V$	8.5×10^{-10}
$PO3 - PO4$	$1.559 \mu V$	$-1.813 \mu V$	3.3×10^{-16}
$O1 - O2$	$0.868 \mu V$	$-0.902 \mu V$	1.2×10^{-4}

mean value of the voltage difference between pairs of contralateral and ipsilateral electrodes in the time interval 280–380 ms after stimulus onset. Since the peak amplitudes did not follow a Gaussian distribution, as assessed by a Lilliefors test for normality, we report the medians of these amplitudes across all participants and trials in Table 4.2 for electrode differences $PO7 - PO8$, $P7 - P8$, $PO3 - PO4$ and $O1 - O2$. The table also reports the p values obtained from the (non-parametric) Kruskal-Wallis test applied to these data. As one can see, the voltage asymmetries documented in these channels (where the N2pc is typically found) are highly statistically significant.

4.3.1.1 Influence of the Presentation Rate on the N2pc Component

Having established the presence of the N2pc for the lowest presentation rate, we can now study this ERP at higher presentation rates. Figure 4.5 shows the grand-averages of the *differences* between corresponding contralateral and ipsilateral ERPs (i.e., equivalent to the line labelled as N2pc in Figure 4.3) across all lateral-target epochs from the training set of one of the folds used in cross-validation, for the different levels of difficulty used in the experiment, measured at electrode sites PO7 and PO8. Again, the shape and timing of the N2pc ERPs are consistent with those reported in the literature [Eimer, 1996; Luck, 2012; Luck & Hillyard, 1994b]. However, in this figure we see three interesting effects: (1) the latency of the N2pc (measured as the time when the difference waveform reaches its peak) tends to become shorter as the presentation rate increases, (2) the peak amplitude for the presentation rate of 6 Hz (i.e., level 2) is larger than that of the presentation rate of 5 Hz, and (3) the peak amplitude at a presentation rate of 15 Hz is the smallest of the four tested.

Generally, one would not expect to see any differences between voltages in the two brain hemispheres during the baseline period preceding a trial (and this is why the plots in Figure 4.5 are virtually zero in the 200 ms preceding the stimulus). However, when stimuli are presented at a high rate, previous lateral targets might “contaminate” the baseline for subsequent trials producing asymmetries in such a period. To check if this could have affected the results, we calculated the percentage of trials where baselines are contaminated by checking, for each presentation rate, how many lateral targets in the dataset are preceded by another target within the previous 400 ms. At the lowest presentation rate (5 Hz),

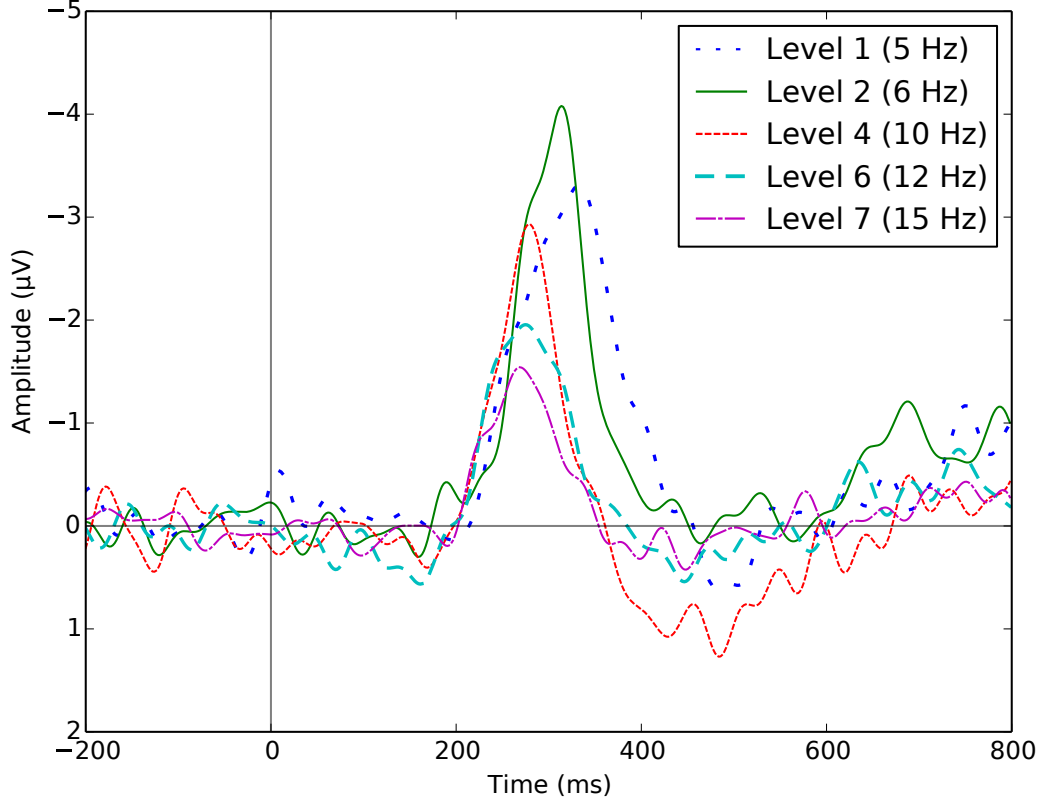


Figure 4.5: Difference plot of the contralateral minus the ipsilateral grand-averages at channels PO7 and PO8 across all lateral targets from the training set for levels with only one type of target.

no trials are contaminated and less than 6% of trials are contaminated at the highest presentation rate of 15 Hz. Other presentation rates fall in between. This indicates that baseline contamination has been minimal in the experiment.

Peak amplitudes were measured for all conditions and tested for statistical differences using a Mann-Whitney U test¹. Results are reported in Table 4.3. This table shows that peak amplitudes for the presentation rates above 10 Hz are significantly smaller than for lower rates. There were no statistical differences

¹The Mann-Whitney U test was used as plotting the data in preliminary tests revealed that peak amplitudes do not follow a Gaussian distribution. This was further corroborated by means of a Lilliefors test for normality that confirmed these results.

Table 4.3: Results of a one-sided Mann-Whitney U test with Bonferroni correction comparing peak amplitudes of the N2pc for different presentation rates (in levels with only one type of target). P values below 0.05 are statistically significant.

	5 Hz	6 Hz	10 Hz	12 Hz	15 Hz
N2pc peak	-2.51 μV	-2.45 μV	-1.12 μV	-0.76 μV	-0.63 μV
5 Hz	–	1	1.9×10^{-11}	1.2×10^{-14}	8.9×10^{-20}
6 Hz	–	–	6.7×10^{-8}	1.8×10^{-10}	1.0×10^{-14}
10 Hz	–	–	–	1	0.09
12 Hz	–	–	–	–	0.86

in peak amplitudes between the presentation rates of 5 and 6 Hz, but the N2pc ERPs that are elicited at these two rates are significantly larger than that at any other presentation rate.

There are at least three possible reasons for these rate-related changes, and they are not mutually exclusive: (1) the target detection task is harder for participants at high presentation rates due to the shorter duration of the target stimuli; (2) the average temporal distance between consecutive targets decreases as the stimulation rate increases, causing some targets to fall within a possible “refractory period” for the N2pc, such as those associated with repetition blindness and the attentional blink [Einhäuser *et al.*, 2007; Jolicœur *et al.*, 2006a; Kanwisher, 1987; Shapiro *et al.*, 1994]; and (3) the choice of experimental design, due to which participants might have been too tired at the end of the experiment, consequently missing more targets than in previous difficulty levels. We elaborate on these factors below.

Task Difficulty We saw in Chapter 3 that there is evidence of the increase in task difficulty as a result of higher presentation rates, resulting in targets being missed. Since participants did not have time to foveate to targets (especially

for the last levels of difficulty), it is possible that those positioned laterally were missed more frequently than those presented in the centre of the screen. Because grand averages do not take into account which lateral targets were seen and which were not, the amplitude of the N2pc component for the high difficulty levels might have been artificially reduced due to the high percentage of missed targets.

Refractory Period of the N2pc Repetition blindness and the attentional blink have been shown to play a role in other ERP-based BCIs [Cinel *et al.*, 2004]. These phenomena manifest themselves as a participant missing a target when the separation from a previous target is less than 500 ms. To test whether some form of refractory period was influencing the ERP amplitudes, the epochs were divided and analysed on the basis of the number of non-targets separating two targets. Figure 4.6 shows grand averages of the N2pc (again, plotted as the contralateral minus the ipsilateral waveforms and using an inverted ordinate axis), for targets that are 2–3, 4–5, 6–7, 8–9, 10–11 and 12–13 stimuli away from the previous target, for the presentation rate of 12 Hz. There are 63, 126, 135, 198, 315 and 180 epochs of each kind, respectively.

At a presentation rate of 12 Hz, the N2pc ERPs associated with well-separated targets (e.g., the line labelled as 10–11) are significantly bigger than the N2pc’s for poorly separated targets, i.e., line 2–3 in the figure. Indeed, the $p < 0.05$ for a one-sided Mann-Whitney U test comparing peak amplitudes of lateral targets that are separated by less than 300 ms — i.e. those labelled as “2–3” in Figure 4.6 — vs the rest, for all samples in the interval 264–307 ms¹. This suggests that refractory phenomena like repetition blindness and the attentional blink may be

¹Note that the Mann-Whitney U test allows comparisons between samples of different sizes.

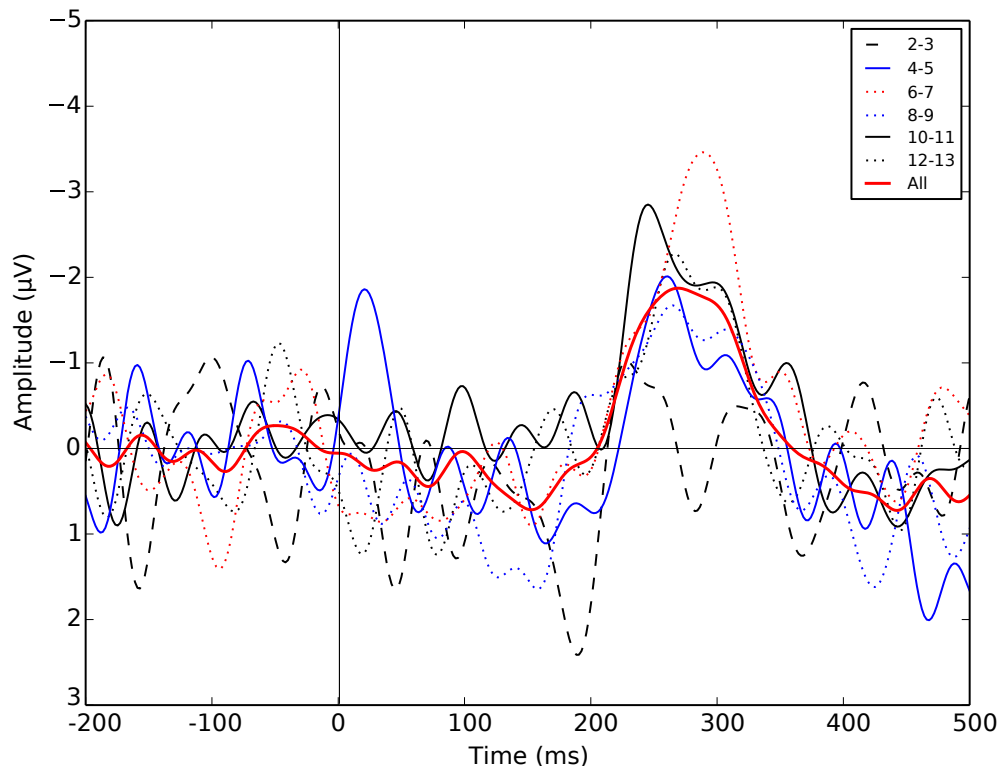


Figure 4.6: Difference plot of the contralateral minus the ipsilateral grand-averages at channels PO7 and PO8 for targets that are within 2–3, 4–5, 6–7, 8–9, 10–11 and 12–13 stimuli away from the previous target, and the grand average (labelled as “all”) across all lateral trials, for a presentation rate of 12 Hz.

partially responsible for the presentation-rate modulations of the N2pc observed.

Experimental Paradigm Finally, the differences in N2pc amplitudes and latencies could partly be attributed to tiredness and learning effects. This is a possibility as the order of the conditions across subjects was not randomised. Randomisation was excluded after receiving early feedback that suggested that participants with no previous experience of high-speed RSVP protocols (such as the cohort of these experiments) found it exceptionally taxing to start with the 10 Hz or 12 Hz conditions. After the standard practice sessions, participants

were able to do reasonably well at the lowest presentation rate of 5 Hz, although in the early blocks many still lamented that the presentation rate was too fast. However, they progressively adjusted and later could cope with increases in the presentation rate. Since the main purpose of the study was to prove the concept that collaborative BCIs can significantly improve the results obtained with single-user BCIs, not to establish whether they are best used at 5, 6 or 10 Hz, we felt that this was a reasonable compromise. However, this design decision implies that the possibility that some of the observed N2pc differences are associated with presentation order effects cannot be excluded.

4.3.1.2 Influence of the Number of Targets on the N2pc Component

We have seen that there are modulations in the EEG that appear as a response to lateral targets in pictures presented in streams at high presentation rates. However, the analysis until now has focused on those levels that contained only one type of target. The experiment included two extra levels (levels 3 and 5, at 6 and 10 Hz respectively) which contained three types of targets. These two levels were added as counterparts to levels 2 and 4, which, at the same presentation rates, respectively, included only one type of airplane. Thus, they may help us understand the differences in the evoked ERPs when participants are asked to look for airplanes vs when they look for a specific type of airplane.

Figure 4.7 reports the contralateral minus ipsilateral grand averages comparing conditions with one vs multiple targets while keeping the presentation rate constant. Shaded areas represent intervals where the two conditions are significantly different (i.e., p value of a one-sided Kruskal-Wallis test was <0.05). The figure shows that the reduced peak amplitudes for levels with several types of

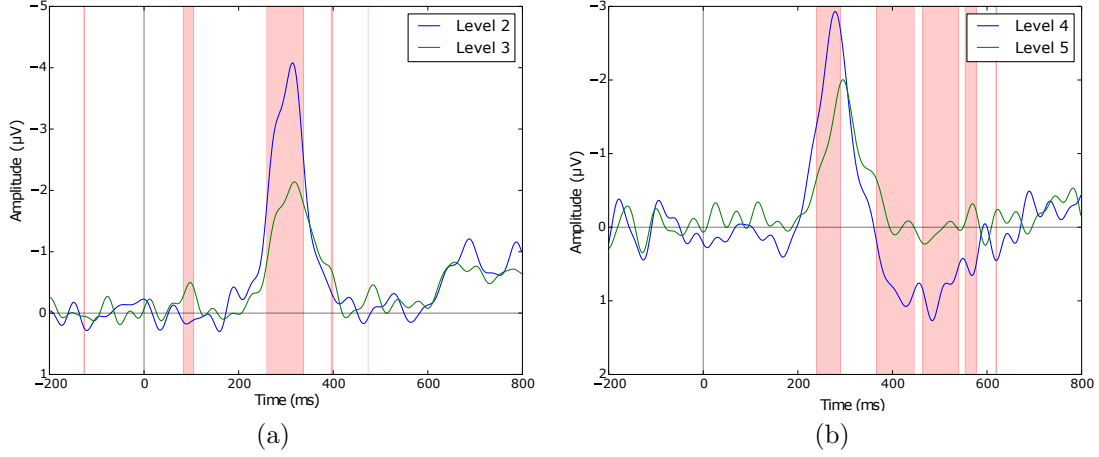


Figure 4.7: Grand averaged contralateral minus ipsilateral differences at channels PO7 and PO8 for levels with one (i.e., levels 2 and 4) vs several (i.e., levels 3 and 5) types of targets at presentation speeds of (a) 6 Hz and (b) 10 Hz. Shaded areas represent time intervals where the two conditions are significantly different.

targets are statistically significant. However, for the higher presentation rate, depicted in Figure 4.7(b), significant differences can also be seen on the tail of the ERP, and not only on the amplitude of the peak.

If again we measure peak amplitudes as the mean amplitude of the ERP in the time interval 280–380 ms after the onset of a lateral target, a Mann-Whitney U test comparing peak amplitudes across levels 2–3 resulted in $p = 1.3 \times 10^{-4}$. However, when comparing peak amplitudes between levels 4–5, the peak amplitude was not significantly different ($p = 0.5$). Thus, the differences highlighted in Figure 4.7 are most likely due to the difference in peak timing, rather than to the amplitude of the peak itself.

Finally, we should note that the peak amplitudes for levels 3 and 5 are not statistically significantly different from each other ($p = 0.08$), while the differences in N2pc peak amplitudes between levels 2 and 4 were significant ($p = 6.7 \times 10^{-8}$). Thus, even though there is a decrease in peak amplitude when moving from a

specific airplane template to a more generic class (i.e., one to three targets), higher speeds within the latter condition do not seem to affect the amplitude of this ERP.

4.3.2 Single-User BCI

The ERP analysis above highlighted the existence of marked asymmetries in the posterior-occipital lateral regions of the scalp in the interval 200–400 ms after stimulus presentation when comparing the ERPs generated by lateral targets at different presentation rates. Therefore, the four electrode differences listed in Section 4.2.3.1 when computed in this interval should allow a classifier to distinguish between LVF and RVF lateral targets.

The results for the single-user BCI approach (described in Section 4.2.3.1) when distinguishing between left and right targets are summarised in Figure 4.8.

As shown in the figure, there are clear large performance variations across participants, evidenced by the range of AUCs covered by the whiskers of the boxplots for each level. Moreover, it can be seen from the figure that the distributions of AUCs tend to be skewed for most presentation rates, although at a lower degree than those of the target detection system (see Figure 3.6)¹.

Despite these large variations within the levels, the median AUCs are reasonably high, which is very encouraging considering the small scalp regions where the N2pc can be detected and its small amplitude. So, overall, classification results indicate that the N2pc can reliably be detected in the conditions of the

¹Since means tend to be affected by outliers and asymmetries in distributions, Medians were used throughout this and the following sections, which give a more robust measure of central tendency in skewed distributions. For this reason, non-parametric statistical tests were also used throughout the chapter.

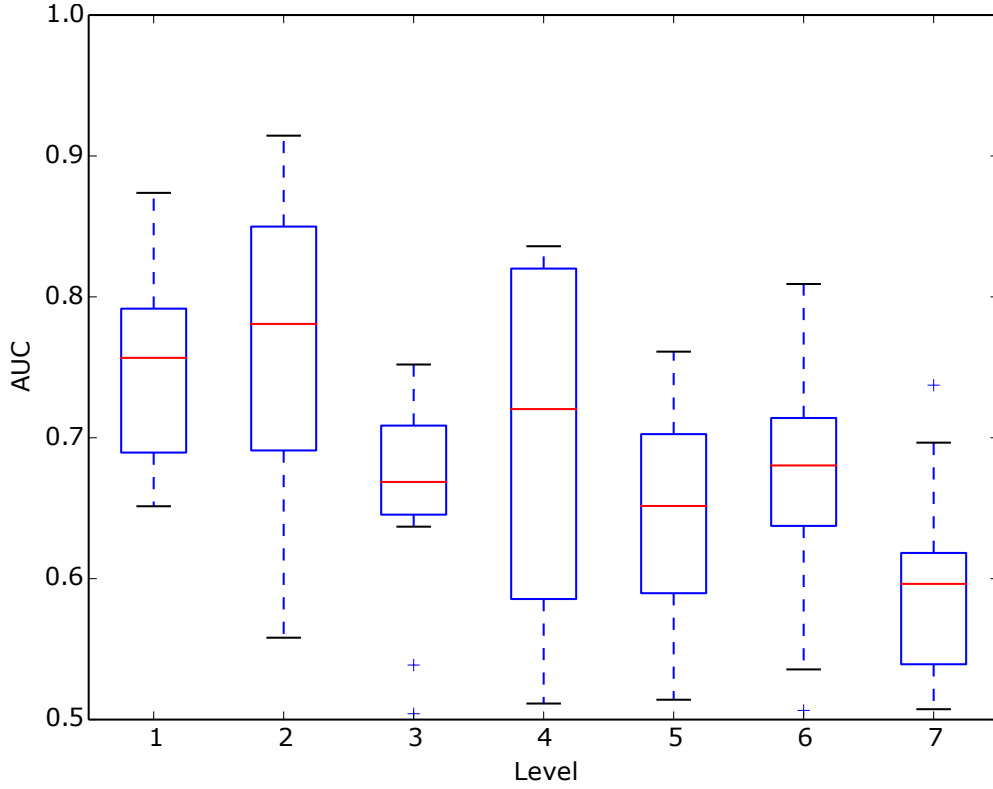


Figure 4.8: Distributions of mean AUCs across the cross-validation folds for single-user BCI across all levels of the experiment.

experiment. Consistently with the ERP analysis, performance drops significantly when moving from level 2 to level 3 ($p = 0.01$). This was expected, as the peak of the difference between contralateral and ipsilateral electrode sites is significantly lower in that condition (as was reported in Figure 4.7). Also consistent with the previous ERP analysis, the decrease in performance that can be observed between levels 4 and 5 is not statistically significant ($p = 0.2$).

It is worth noting, however, that performance for most participants is well above that of a random classifier (i.e., $AUC = 0.5$) and that the top quartile have $AUCs \geq 0.7$ even at a presentation rate of 12 Hz (i.e., level 6). This suggests that with a suitable participant selection process, the BCI could also be successfully

operated even at that rate.

4.3.3 Handedness, Target Localisation and the N2pc

After initially trying to predict target localisation and obtaining promising results, we noticed some differences in the ERPs produced by left- and right-handed participants. This section contains the results of some additional analyses that we performed after noticing this effect.

4.3.3.1 Handedness Classification

Figure 4.9 shows the contralateral minus ipsilateral grand averages across lateral targets in the training set from one of the cross-validation folds as described in Section 4.2.3.3 for LH and RH participants. This figure also reports the p values from a one-sided Mann-Whitney U test comparing ERP amplitudes over time. As shown, there are highly significant differences between the N2pc in LH and RH participants (i.e., all those where the values of p are above the 5% horizontal line), especially in the tail of the ERP in the time window 280–400 ms after the onset of the stimuli.

Lateral targets were then separated according to their position on the screen. Figure 4.10 shows the N2pc grand averages obtained for left- and right-handed participants, separated for LVF (Figure 4.10(a)) and RVF targets (Figure 4.10(b)). Even though there are significant differences in both LVF and RVF targets as a result of the handedness of the participant, the most pronounced differences are found in the N2pc component for targets located on the right side of the screen. For this reason, it was decided to use only epochs from such targets for the classification of participants based on handedness.

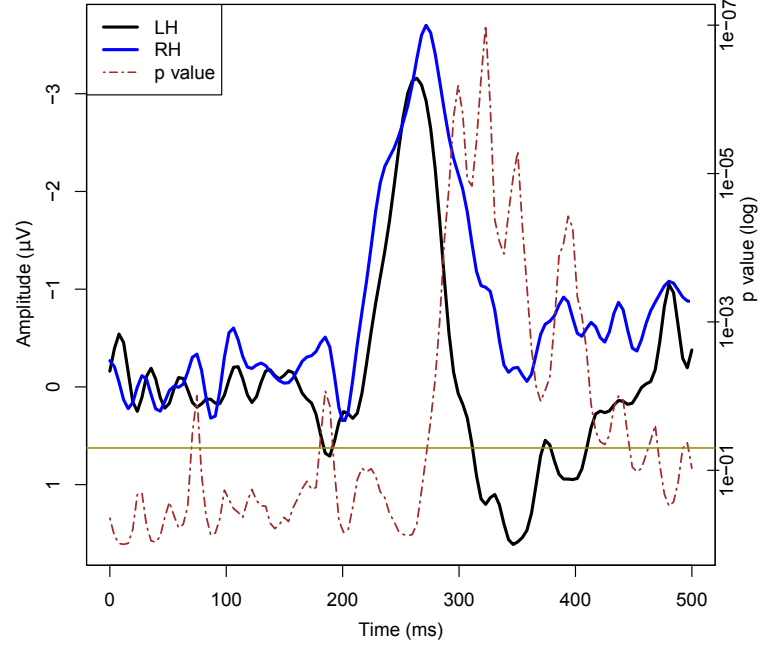


Figure 4.9: Contralateral minus ipsilateral stimulus-locked grand averages for LH and RH participants across lateral targets from the training set of one fold at the presentation rate of 5 Hz, and p values from a one-sided Mann-Whitney U test comparing both conditions.

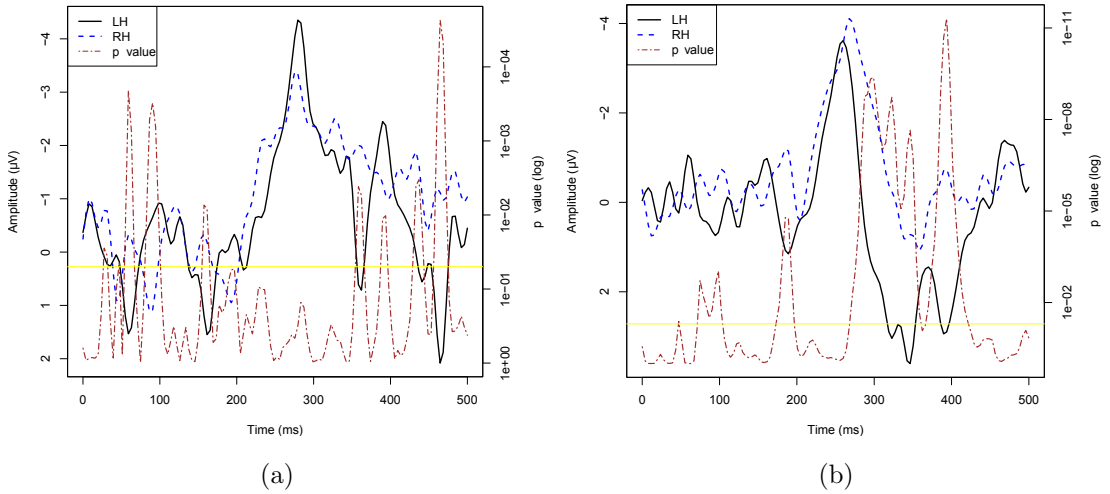


Figure 4.10: Contralateral minus ipsilateral stimulus-locked grand averages for LH and RH participants across (a) LVF and (b) RVF targets from the training set of one fold at the presentation rate of 5 Hz, and p values from a one-sided Mann-Whitney U test comparing both conditions.

Using the methods described in Section 4.2.3.2, and a leave-one-subject-out loop, a classification accuracy of 100% was obtained for the discrimination of participants based on handedness (i.e., LH vs RH) in cross-validation.

As we will discuss later, the differences in the N2pc due to handedness should be taken into account when working with collaborative BCIs, especially if EEG signals from the individuals that compose a group are averaged directly, e.g., as in the SC-cBCI approach.

4.3.3.2 Prediction of the Analogue Horizontal Position of Targets

As hypothesised, for both LH and RH participants the amplitude of the N2pc can also be used to roughly determine the distance of the target from the centre of the picture. As an example, Figure 4.11 shows the coordinates of the target output by the PSO-optimised linear predictor (trained as explained in Section 4.2.3.3) vs the actual coordinates of the target for the group of 5 left-handed participants. The correlation between the predictor's output and the actual x-coordinate of the target on the *test set* for the LH group is $\rho=0.42$ (and $\rho=0.39$ for the RH group). In all cases, the 8 features selected by the PSO represent the amplitude of the N2pc in electrode differences between PO7-PO8, P7-P8 and PO3-PO4, but not on O1-O2.

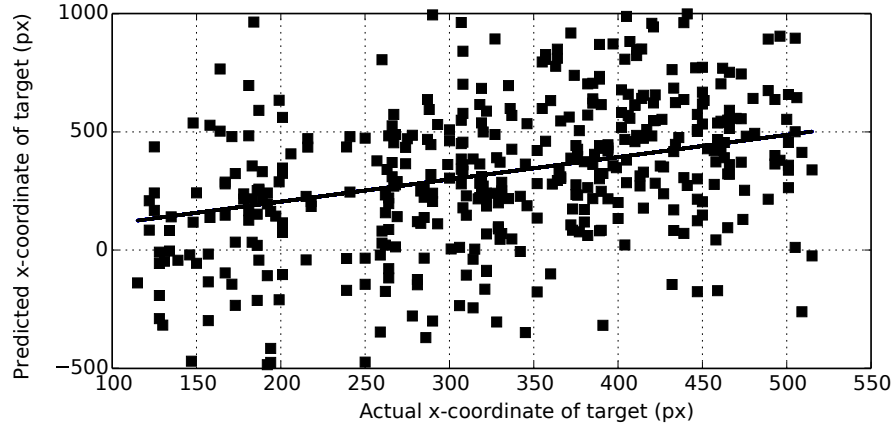


Figure 4.11: Predicted x-coordinate for the target vs actual target position (in pixels) for all targets in the test set, using only LH participants. The regression line is also shown.

4.3.3.3 Influence of the N2pc on Target Detection

As described in Section 4.2.3.5, several different combinations of electrodes were tested for the classification of T vs NT trials in order to explore the influence of lateralised information when discriminating between targets and non-targets.

The first combination (E_{28}) was formed of centro-parietal electrodes where the P300 ERP is most prominent; the second one did not include those electrodes that were used for the detection of the N2pc (E_{20}); and finally a third combination (E_{24}) used the features from the electrodes in combination E_{20} plus the 4 pairs of electrode differences used for N2pc detection.

Table 4.4 reports the median AUC values obtained across all participants for each level and electrode combination. A one-sided paired Wilcoxon rank test comparing the participant-by-participant results revealed that the small difference in medians for combinations E_{28} and E_{20} is statistically significant (p value= 4.4×10^{-4} after Bonferroni correction). This confirms that there is a consistent, albeit

Table 4.4: Median AUC values across all participants for each difficulty level and combination of electrodes, for T vs NT classification.

	<i>Level 1</i>	<i>Level 2</i>	<i>Level 3</i>	<i>Level 4</i>	<i>Level 5</i>	<i>Level 6</i>	<i>Level 7</i>
E_{20}	0.86	0.84	0.70	0.76	0.58	0.65	0.58
E_{24}	0.86	0.86	0.68	0.77	0.58	0.67	0.58
E_{28}	0.87	0.87	0.69	0.79	0.57	0.70	0.61

small, advantage in integrating channels where the N2pc is typically present with those where P300s are most prominent for the purpose of detecting targets. The difference in medians between E_{24} and E_{28} is also significant (p value= 7×10^{-4} , after Bonferroni correction). However, comparisons between E_{20} and E_{24} revealed no statistically significant differences (p value=0.29). This indicates that the hemispheric asymmetries of the N2pc do not particularly help target detection.

In addition to the AUC comparisons, we estimated the probability density functions (pdfs) for the SVM output scores obtained for T vs NT classification using the E_{28} and E_{20} electrode combinations. Since we wanted to distinguish between the cases where an N2pc is expected from those where it is unlikely to be elicited, we computed separate (conditional) pdfs for lateral and central targets as well as the (total) pdf for all non-targets. Figure 4.12 shows the results. To provide a common reference and make it possible to appreciate relative differences between classes, the pdfs were normalised by subtracting the mean of the non-target scores and were then scaled by the standard deviation of the same class (so the non-target pdf has zero mean and unitary standard deviation).

As shown in the figure, the pdfs for targets and non-targets are reasonably well separated (as also highlighted by our earlier AUC analysis). However, both for E_{28} and E_{20} we see that the pdfs for central and lateral targets differ to some degree, with the central targets achieving higher scores than the lateral

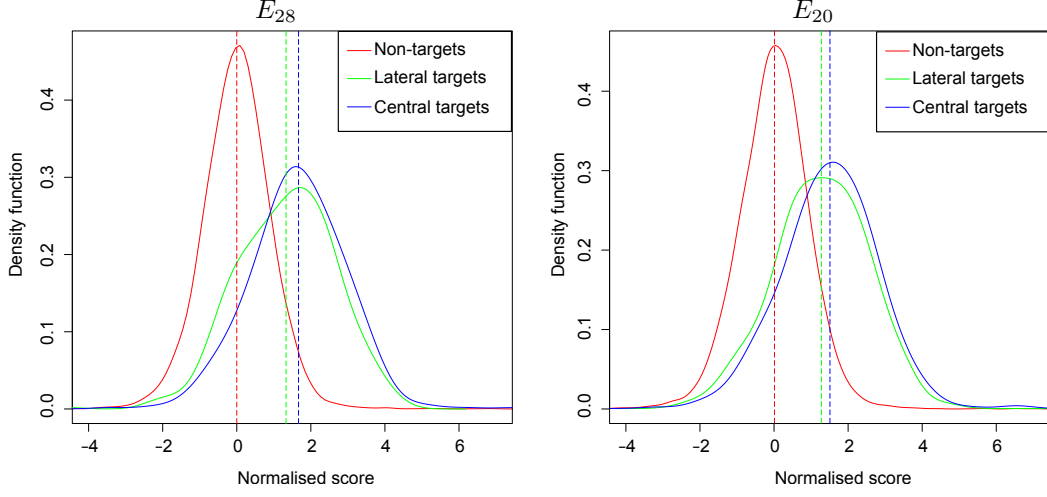


Figure 4.12: Probability density functions of the SVM normalised scores in the T vs NT classification for non-targets, lateral targets and central targets using combinations E_{28} (left) and E_{20} (right) electrodes, computed via R's Gaussian-kernel-based density estimator. Vertical lines represent the mean of each distribution.

targets (approximately 1.67 vs 1.33 for E_{28} and 1.50 vs 1.27 for E_{20} , respectively). To verify if these differences are statistically significant, we performed a one-sided Mann-Whitney test to compare the medians of these distributions across participants. The results show that the score shifts of lateral targets with respect to central ones are highly statistically significant for combinations E_{28} (p value = 3.9×10^{-5}), E_{20} (p value = 0.001) and E_{24} (p value = 0.002). This observed shift in the distributions for lateral targets is likely to be a manifestation of P300s being different for lateral targets than for central ones [Dias & Parra, 2011; Marathe *et al.*, 2016].

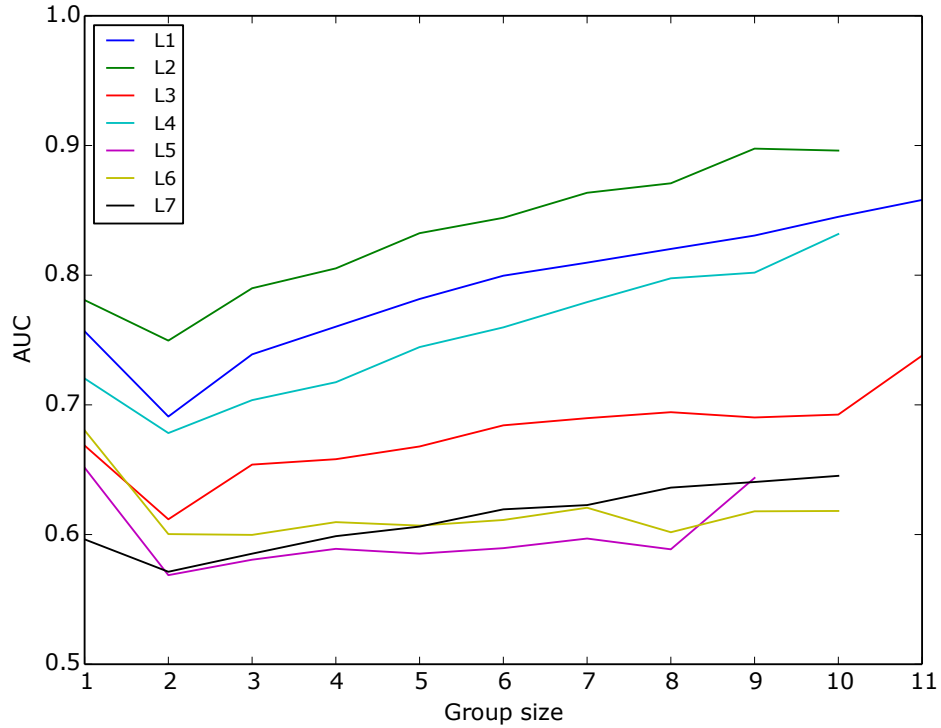


Figure 4.13: Performance of the SC-cBCI in terms of the median AUC for each group size and level.

4.3.4 Collaborative Classification

We saw in Chapter 3 that combining evidence from several users can help improve the performance of a BCI for target detection. This section reports an equivalent analysis applied to target localisation by means of the N2pc.

Figures 4.13–4.15 show the average AUC values across all possible groups for each group size and level, for the SC-cBCI, MC-cBCI and LDA-cBCI methods, respectively. Again, it can be observed that bigger groups lead to better performance for all methods. Moreover, the approaches based on a group classifier (i.e., MC-cBCI and LDA-cBCI) outperform the simple average of groups' signals.

One striking feature that did not appear on Chapter 3 for the SC-cBCI case is the somehow surprising decrease of performance for groups of size two. This effect

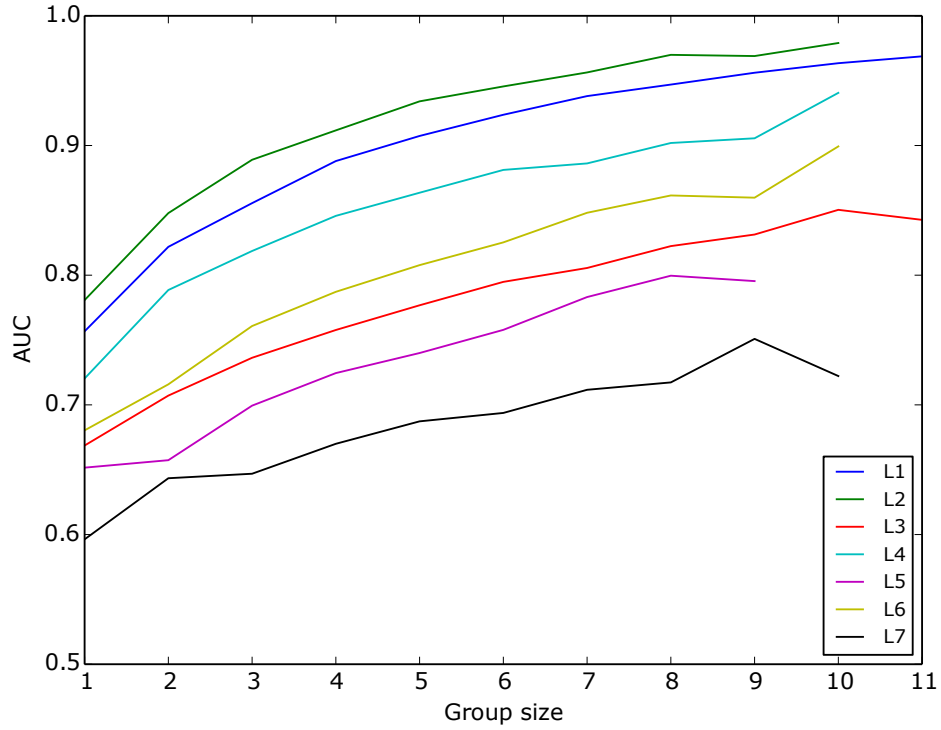


Figure 4.14: Performance of the MC-cBCI in terms of the median AUC for each group size and level.

will be studied in Chapter 6, where it will be proposed that it might be due to the distributions of AUCs for the single-user target localisation BCIs. Moreover, it will also be studied whether this drop can be reduced by selecting participants when forming the groups for the cBCI.

Table 4.5 summarises the results of pairwise comparisons between the three methods. Out of the 192 possible one-sided comparisons, 130 were statistically significant (after Bonferroni correction). The full table of results and p values can be found in Appendix B.

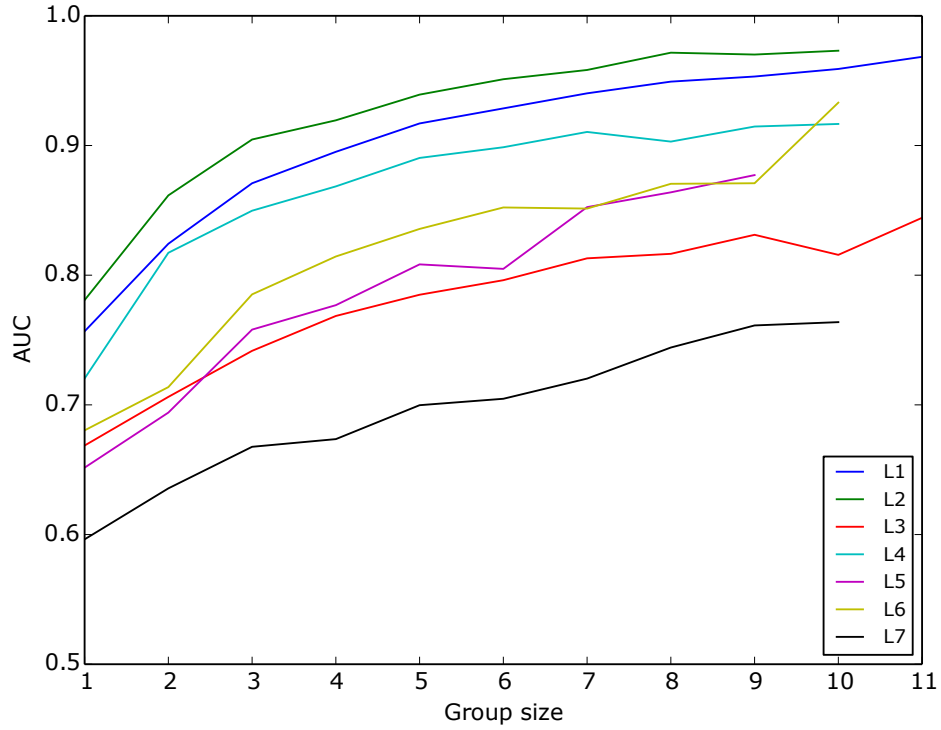


Figure 4.15: Performance of the LDA-cBCI in terms of the median AUC for each group size and level.

Table 4.5: Summary of comparisons between the 3 modes of creating cBCIs: LDA-cBCI (“L”), MC-cBCI (“M”) and SC-cBCI (“S”). The symbol > indicates that the method on the left is significantly better than the one of the right. The methods on both sides of symbol = are not significantly different. N/A values represent those cases in which there were not enough samples to perform the test.

<i>Group size</i>	<i>Level 1</i>	<i>Level 2</i>	<i>Level 3</i>	<i>Level 4</i>	<i>Level 5</i>	<i>Level 6</i>	<i>Level 7</i>
2	L=M>S	L=M>S	L=M>S	L>M>S	L=M>S	L=M>S	L=M>S
3	L>M>S	L>M>S	L=M>S	L>M>S	L>M>S	L>M>S	L=M>S
4	L>M>S	L>M>S	L>M>S	L>M>S	L>S>M	L>M>S	L=M>S
5	L>M>S	L>M>S	L>M>S	L>M>S	L>S>M	L>M>S	L>M>S
6	L>M>S	L>M>S	L=M>S	L>M>S	L>M>S	L>M>S	L>M>S
7	L=M>S	L>M>S	L>M>S	L>M>S	L>M>S	L=M>S	L=M>S
8	L=M>S	L>S	M>L>S	L=M>S	N/A	L=M>S	L>M>S
9	L=M>S	N/A	L=M>S	N/A	N/A	N/A	N/A

4.4 Discussion

One of the objectives of this chapter was to study whether the N2pc ERP was present during the rapid presentation of real-world images. The ERP analysis revealed the presence of the N2pc, and that it had a greater latency than had previously been reported in the literature. In part, this may be due to the greater complexity of the stimuli used in this study in comparison to the simple stimuli traditionally used in the literature. However, we suspect that the stimulus presentation technique used here (i.e., RSVP) is the prominent reason for the greater latency. In typical N2pc-evoking experiments, participants are shown an array of objects or symbols either for a short amount of time (usually <300 ms) with a generous inter-stimulus interval (>1.5 s), or until they find the target. However, in the paradigm used here, images follow each other very quickly, and, thus, a target image is immediately followed by one or more non-targets. We hypothesise that these effectively act as masks for the target picture, thus resulting in a significant increase in the cognitive load of the task and diverting attentional resources away from it.

Moreover, the N2pc components evoked using this paradigm change in both amplitude and latency as the presentation rate is varied. Some potential sources for such variations were also analysed, and we were able to confirm that these are (at least in part) due to a “refractory period” behaviour of the visual system.

These effects are not only present for the N2pc component: numerous studies have researched variations of the P300 ERPs using different types of stimuli in BCIs, more specifically using Donchin’s speller. For example, P300s are known to be significantly deformed at high presentation rates, due to several contribution

factors, such as the brain not getting back to a rest state after the ERP (because there is one or more other P300s overlapping with it), the refractory period of the P300 ERP, the SSVEPs generated by the stimulation (regardless of whether or not participants are carrying out any mental tasks) and the masking effect of new incoming stimuli [Citi *et al.*, 2009, 2010; Hill *et al.*, 2008; Martens *et al.*, 2009; Salvaris *et al.*, 2012]. All these factors (and many others) might be involved in the variations in the N2pc ERPs elicited.

If we now look at the BCI's performance, we see that not only is the N2pc evoked in the conditions of this experiment, but results also indicate that it can reliably be detected even at high presentation rates. For instance, single-user BCIs were able to obtain a median AUC value of 0.76 for single-trial LVF vs RVF classification based on the detection of this ERP using only 4 electrode differences and a very short time window of approximately 200 ms for the presentation rate of 5 Hz. At high presentation frequencies, results for the single-trial single-user BCI indicate that the N2pc can still be detected for presentation rates of up to 12 Hz. Of course, performance drops with increasing speeds, although it is well above chance levels for most participants and the drop from the easiest to the most difficult level is not as pronounced as in the target detection system of Chapter 3. Moreover, the drop in performance when moving from levels with one target template to levels with three (i.e., levels 2–3 and 4–5) was not as marked as in the P300 detection task.

By using different methods of combining classifiers' outputs, we also saw that cBCIs significantly outperform single-user BCIs in the left vs right classification task. The best method for combining evidence from participants to form groups was based on an LDA classifier. This approach was, in every case, either on a

par or significantly superior to the other methods used in this chapter.

In relation to handedness, it was shown that the shape of the N2pc component that is elicited in response to targets that appear on the left and right side of the screen differs for RH and LH participants. Such differences, which occur both in the timing of the peak and in the tail of the ERP, had not been previously reported in the literature, and are large enough to allow a classifier to discriminate between left- and right-handed participants. However, since the volunteers did not perform a handedness test, these results should be taken as a preliminary result and should be further studied in the future. Although there is previous evidence of the dependence between handedness and memory [Lyle *et al.*, 2008], handedness and brain morphology [Habib *et al.*, 1991; Witelson, 1985] and EEG signals being different depending on handedness [Nielsen *et al.*, 1990], to the best of our knowledge, these differences have not been exploited to assess the handedness (or preferred hand) of a person¹.

Within each handedness group, it was revealed that the N2pc can not only be used to distinguish between LVF and RVF targets, but it can also tell to what degree a target is lateral. Taking into account the differences in the N2pc due to handedness, participants were separated into a left- and a right-handed groups, and a linear regressor was trained to predict the x-coordinate of the target. The correlations obtained between the real and the predicted coordinates, although promising, are barely enough to get a sense of the approximate location of the target.

¹The only article that we are aware of that touches upon this subject is [Ng & Leong, 2014]. However, there are a number of flaws in the methodologies and inconsistencies in the results, including setting the threshold for classification based on the test data, contradictions in whether or not the participants from the second experiment (whose results are never reported) were left-handed, and forgetting to include the results claimed in (their) Section 4.5.

4.5 Conclusions

This work attempted to locate the position of targets on the x-coordinate of images presented through RSVP. To the best of our knowledge, this is a novel application that had not been attempted before. We also found that it is possible, to a degree, to determine how lateral the target is, although further improvements of the work presented here are needed. This avenue for future work will be explored in Chapter 5.

As a byproduct of our research, we discovered a relationship between the N2pc and handedness, which is evident enough to discriminate participants using the ERPs evoked in response to lateralised targets.

The findings of this chapter suggest that there could be ways of exploiting handedness and lateralisation (as emphasised by the N2pc) to build even better performing integrated T vs NT *and* LVF vs RVF classification systems. Even though we will not further pursue the former topic (i.e., taking into account the handedness of the participant when creating groups, we believe that a more detailed study of the implications of handedness on target detection should be done in the future. The latter topic (i.e, concatenating the target detection and target localisation systems) will be the focus of study for Chapter 5.

Moreover, further radical improvements in both the target detection and the target localisation systems can also be obtained by selecting the observers from which the neural evidence is integrated, e.g., by means of member selection when forming the groups. This research avenue will be explored in Chapter 6.

Chapter 5

Concatenating Detection and Localisation Systems

We have seen that both the P300 and the N2pc contain information useful for the detection and localisation of targets. We will now study different ways of concatenating both systems in order to explore whether they can help each other and to increase the information transfer rate of the BCIs presented so far.

5.1 Introduction

Previous chapters showed that both the P300 and the N2pc ERPs contain valuable information that can allow a system to detect (Chapter 3) and locate (Chapter 4) targets or events of interest in rapid streams of pictures. This chapter goes one step further and combines these two systems in order to obtain a full BCI that can be used in a real-world setup to detect targets while at the same time providing their approximate location, thereby reducing the complexity and time

required by further analysis (e.g., by a second observer tasked with confirming the detections of the first).

This chapter is organised as follows: Section 5.2 gives a quick reference to the data collection and preprocessing methods used in previous chapters and which are also used in this work. It also presents three approaches to combine the P300-based and the N2pc-based systems: the sequential N–P system, the sequential P–N system and the sequential regression system. The results are presented individually for each of the approaches in Section 5.3, and discussed in Section 5.4. Finally, Section 5.5 concludes the chapter and presents some avenues for future work.

5.2 Methods

Most of the basic methods and techniques used in this chapter are either the same or very similar to those described in the target detection and target localisation chapters. This section will give the reader a brief summary of the main aspects of those systems (with pointers to the corresponding sections where they were described in detail) and proceed to the description of how the detection and localisation systems were combined, in both the individual and, where applicable, the collaborative BCI cases.

5.2.1 Participants and Setup

The participants and experimental protocol were the same as in Chapters 3 and 4. Data were filtered between 0.15–28 Hz and downsampled to 64 Hz. Epochs were extracted for each trial (i.e., each image) and baseline-referenced to the mean of

the 200 ms interval preceding the onset of a trial.

5.2.2 Feature Extraction

N2pc and P300 features were extracted from every trial (regardless of whether it contained a target or not). The N2pc features were extracted as described in Section 4.2.3.1, and the P300 features were extracted using the E_{20} combination of electrodes (shown in blue in Figure 4.2) and the interval 300–600 ms after the onset of the stimulus, as described in Sections 3.2.3 and 4.2.3.5.

Three ways of joining the target detector and the target locator systems were tested:

- A system that used the score of the target locator as an extra input feature to the P300 detector (called the *sequential N–P system* hereafter and described in Section 5.2.3).
- A system that used the score from the P300 system as an extra feature for the N2pc classifier (*sequential P–N system*, described in Section 5.2.4).
- A system that changed the N2pc SVM-based classifier of the sequential P–N system with a neural network trained to perform regression based on the input features (the *sequential regression system*, described in Section 5.2.5).

In order to measure the performance of each single-user system, the dataset from each participant was divided into 10 stratified folds using four classes: (1) non-targets (NT), (2) central targets (CT, i.e., those that contain a target, but did not fit the criteria to be considered lateral targets), (3) LVF targets, and

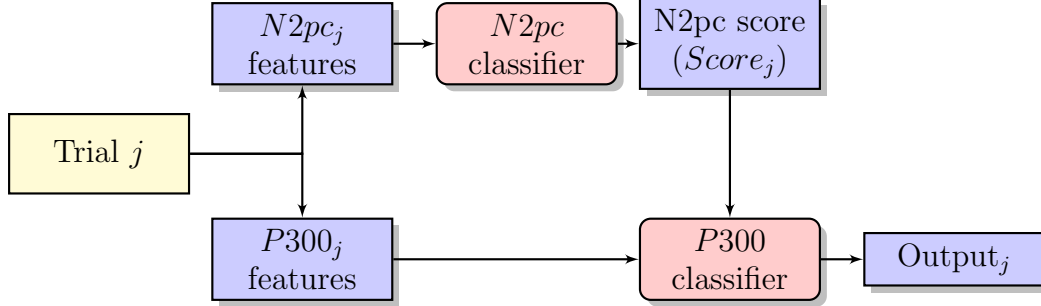


Figure 5.1: Pipeline of the sequential N–P system.

(4) RVF targets. Thanks to the stratification, every training and test fold had approximately the same proportion of trials each class as the full dataset.

5.2.3 Sequential N–P System

Figure 5.1 shows the processing pipeline for the classification of a new trial in the sequential N–P approach to combining the systems.

As shown in the figure, for each fold, N2pc and P300 features are extracted from each trial. The N2pc features are fed to the N2pc classifier, that has been trained using only the LVF and RVF trials from the training set of the current fold. This score is fed into the P300 classifier together with the P300 features previously extracted. The final output of the system is then given by the P300 classifier.

5.2.3.1 Rationale

We saw in the literature review (Section 2.2.2) that the N2pc ERP is believed to signal a shift of covert attention towards a potentially relevant element that is lateral [Luck, 2012]. Hence, the N2pc classifier used in this approach would act

as an attention shift filter before the target detection task. The decision of using only lateral targets to train the classifier (instead of, for instance, dividing the screen in two and labelling a target as left or right depending on which half of the picture they fall on) is based on the results from Chapter 4, where we saw that the amplitude of the N2pc can give an idea of the degree of laterality of a target. Hence, central targets will have a near-zero score (like non-targets would), which will introduce noise in the training set.

The rationale behind the sequential N–P approach is that if there was a lateral element in an image that captured the viewer’s attention (thus resulting in a large N2pc score), it is more likely to have been a target. If the hypothesis is correct, large values of the N2pc score might be helpful to train the P300 classifier to better distinguish targets, since we also saw in Section 4.3.3.3 that the scores from the target detection system are different for lateral than for central targets.

For all the reasons above, it was hypothesised that the detection of the targets could be improved by adding the score from the N2pc (or attention filter) classifier, hence increasing the ITR of the target detection system.

5.2.3.2 System Training

As indicated above, for each fold, the attention shift classifier was trained using the N2pc features from the LVF and RVF targets from the training set. As in Chapter 4, the classifier (an ensemble of two linear SVMs) was trained to distinguish between the two types of lateral targets.

After training the N2pc classifier, all the training epochs (belonging to the four possible classes) were fed to this classifier in order to obtain a score for each one. This score was fed to the P300 classifier together with the P300 feature vector of

the trial, and the P300 classifier was trained to perform T vs NT discrimination using the P300 features and the output score from the N2pc classifier.

In this case, the target (T) class was composed by all instances from the training set of each fold that belonged to the CT, LVF and RVF classes.

5.2.3.3 System Testing

This system was an attempt at trying to improve target detection by taking into account an indicator of whether there was something in an image that caused a shift in the participant's attention.

For this reason, the metrics used to evaluate this system are the AUC and the ITR, both calculated for each participant as the mean value obtained for the test set of each fold, across the 10 stratified cross-validation folds.

5.2.4 Sequential P–N System

Figure 5.2 shows the processing pipeline for the classification of a new trial in the sequential P–N approach to combining the target detection and localisation systems.

In this approach, the order of the target detector and the target locator is reversed with respect to the N–P system presented above. Again, both P300 and N2pc features are extracted from each trial. In this case, however, the P300 classifier is fed the P300 features only (note that the P300 classifier from the sequential N–P system was trained on the features from the P300 and the output score from the N2pc classifier), and its output score is then used to decide whether the image contained a target or not. If the system believes that there was a target, the P300 score from the trial is fed, together with the N2pc feature vector, to the

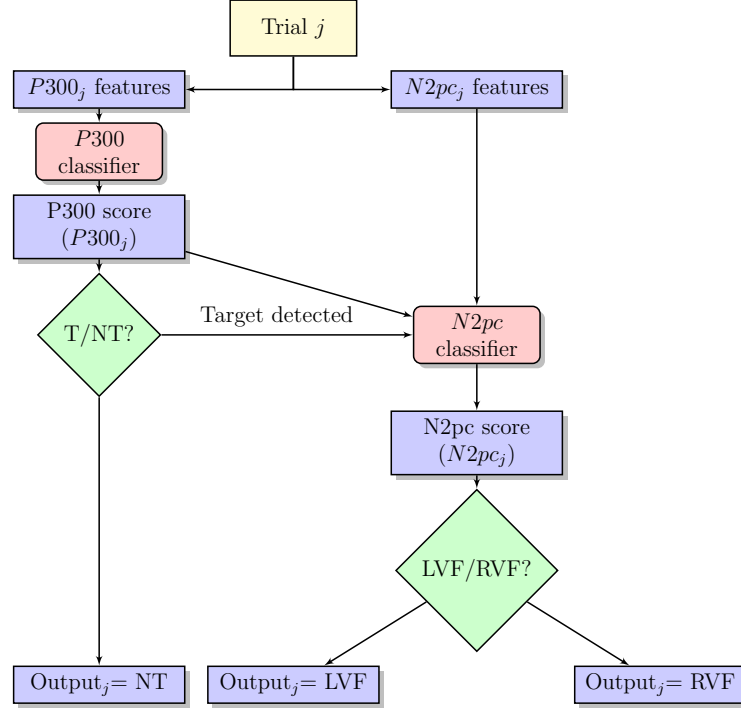


Figure 5.2: Pipeline of the sequential P–N system.

N2pc classifier, which decides whether the target was located on the left or the right hemifield of the picture.

5.2.4.1 Rationale

The sequential P–N system first uses the target detector to establish whether the picture associated with an incoming trial contained a target. If it detects a target, it then tries to delimit the area of the picture where it is located by means of the target location system. This is another logical approach to concatenating the target detection and the target location systems.

In contrast to the sequential N–P system, which aimed at improving the detection of targets by adding the output of the attention shift classifier, the P–N

approach is not expected to improve the performance of the N2pc classifier itself: the target localisation system will now have to work with central targets for which it has not been trained with (see below), so performance might even decrease due to the division of left and right side targets in the middle of the screen, rather than leaving the centre of the image out of the analysis as was done in the previous chapter. However, by increasing the number of classes that the system is able to detect from the binary T/NT to three classes: LVF, RVF and NT, there is a chance that the ITR of the combined system could improve.

As seen in previous chapters (see Section 4.3.3.3), the distributions of raw scores of the P300 classifier differ significantly depending on the location of targets. For this reason, it was decided to feed the raw score of a target as an extra feature to the N2pc classifier, hoping to decrease errors in central targets and help improve the target localisation task¹.

In order to measure the performance of this system, as it will be described below, an extra metric was used in addition to those mentioned previously: the confusion matrix.

5.2.4.2 System Training

The P300 classifier was trained in the usual way: all targets were grouped into only one target class, regardless of location. For each fold, an ensemble of 2 linear SVMs was trained using the P300 features from all trials of the training set divided into targets and non-targets.

¹This extra feature would help if we were interested in the coordinate of the target, but, as explained above, we did not hope to improve the AUC of the N2pc system. Instead, we decided to keep this for compatibility with the sequential regressor system introduced in the following sections

The N2pc classifier was trained using the N2pc features together with the raw P300 score using only the lateral (LVF and RVF) targets of the training set of that fold.

5.2.4.3 System Testing

For each trial from the test set, the output of the P300 classifier was temporarily transformed into a binary decision as to whether the picture contained or not a target. If it did, the original raw score from the classifier was fed, together with the N2pc features for that trial, to the N2pc classifier.

Since no central targets had been given to the N2pc classifier during training, at the time of deciding the location of a central target, in order to measure the performance of the system, a target was considered to be a left (resp. right) target if the x-coordinate of its centroid was in the left (resp. right) half of the picture.

The performance of the system was measured in terms of the mean AUC and ITR across the 10 cross-validation folds for each participant. The AUC was calculated at the output of the N2pc classifier, using only the left and right target classes for comparison with the single-user BCI system from Chapter 4.

However, the confusion matrices and ITRs of the system were calculated taking into account the three classes that can be output by the system: non-target, left-side target and right-side target.

5.2.5 Sequential Regression System

The processing pipeline for the sequential regression system is depicted in Figure 5.3. Although the pipeline of this approach is similar to that of the sequential P–N system, a few differences were introduced when creating this system.

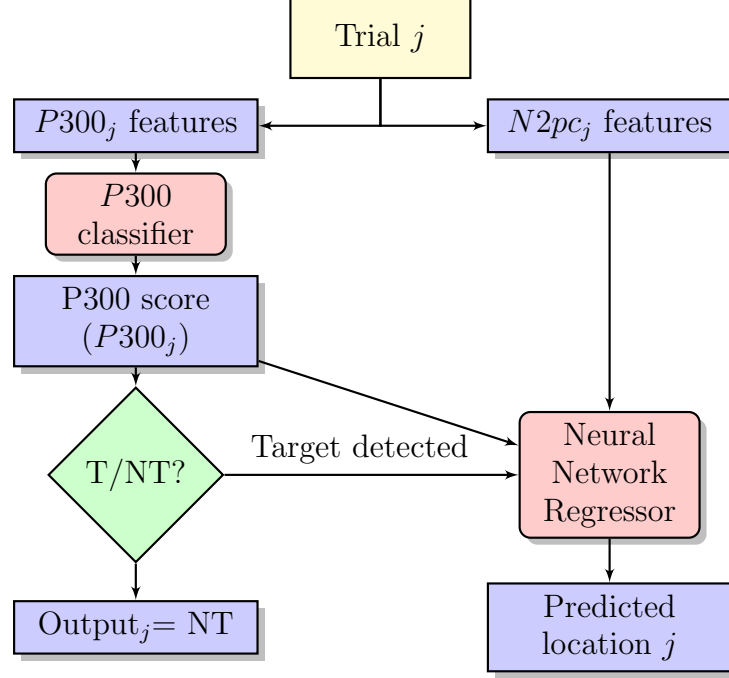


Figure 5.3: Pipeline of the sequential regression system.

5.2.5.1 Rationale

The previous two approaches did not take full advantage of all the knowledge that has been acquired in the previous chapters. In particular, it was shown in Chapter 4 that it is possible to correlate the outputs of a linear model trained by means of a PSO with the position of the targets (Section 4.3.3.2).

This fact is taken into account in the sequential regression system, which keeps the target detection algorithm of the P–N system, but substitutes the N2pc classifier with a neural network that is trained as a regressor.

We decided to use a neural network instead of a standard linear regressor because multivariate regression methods are known to compress their output range in the presence of strong noise in the inputs. In preliminary tests, we

attempted to perform standard linear regression and other regression methods. However, the results were not satisfactory, so we chose to use neural networks as they provide a more advanced approach for this experiment.

5.2.5.2 System Training

The P300 classifier was trained as in Section 5.2.4, using only the P300 features to discriminate between targets (CT, LVF and RVF, grouped into one class) and non-targets (NT). As in the P–N system, if the P300 classifier found a target, its raw score was fed to the N2pc neural network as an extra feature.

Each neural network was trained using the N2pc features extracted from all targets from the training set of each fold, plus their raw output scores from the P300 classifier. The labels used during training were not the discrete classes used until now, but the coordinate of the targets.

Since the PSO system used in the previous chapter (Section 4.3.3.2) was able to obtain reasonable results using only 8 features from the original vector of 56 that is originally extracted for N2pc classification, we decided to use a simple neural network architecture for the regression of target locations based on the N2pc features from the EEG. This decision was also motivated by the fact that we did not have a lot of data available to train the neural network, and we did not want to risk overfitting the system by using too many layers and/or neurons per layer.

Thus, the selected architecture contained a hidden layer whose parameters (number of neurons, n , and type of activation function) were obtained via a cross-validation grid search [Pedregosa *et al.*, 2011] within each of the cross-validation folds. In particular, the grids to be explored by the search were the number of

neurons for the hidden layer, $n \in \{5, 10, 20\}$, and the type of activation function for this layer, which could be tanh (hyperbolic tangent) or sigmoid. The output of the system was given by a linear neuron. The neural network was trained through gradient descent, and the learning rate of the neurons was adjusted using the RMSprop algorithm [Tieleman & Hinton, 2012], which adjusts the learning rate depending on the magnitude of the recent gradients.

5.2.5.3 System Testing

The main goal of this system was to predict whether a given image contains a target and, if so, what its x-coordinate is. For this reason, and considering that the performance of the T vs NT classifier has been studied in previous sections and chapters, the performance analysis of the sequential regression system focused only on the correlation between the actual and the predicted coordinates, which, in turn, is a measure of the predictive power of the neural network.

In addition to this, the changes in the slope of a linear regressor fitted to the outputs of the neural network (predicted coordinate vs real coordinate on the test set) were studied. The correlation coefficient ρ and the regression slope β are related through the following equation:

$$\beta = \rho \times \frac{\sigma_{output}}{\sigma_{input}}$$

where σ_{output} and σ_{input} are, respectively, the standard deviations of the output (i.e., the values obtained from the linear regressor) and the input (in this case, the real x-coordinates of the targets) values. In the case of standardised variables, $\beta=\rho$. However, in the rest of the cases, they give different information about the

strength of the linear relationship between inputs and outputs: the correlation coefficient is independent of the scale of the variables, and gives information about how close they are to a perfect linear relationship; the regression slope is the change in the expected value of the outputs that corresponds to a change of one unit in the inputs.

In the proposed system, given that the inputs are always the same (and correspond to the x-coordinates of the targets in the RSVP experiment), changes in $\frac{\sigma_{output}}{\sigma_{input}}$ will effectively be a reflection of changes in the variance of the outputs.

5.2.5.4 Collaborative Sequential Regression System

As we saw in Chapter 4 (Section 4.3.3.2), a PSO regressor trained with data from multiple participants is able to predict the location of the target in terms of its actual coordinate, rather than just the hemifield. Up to this point, collaborative BCIs have been used to improve the *classification* task, both for target detection and target localisation. However, on this occasion we studied the improvements that can be obtained in the sequential regression system when combining information from multiple users, thus creating a cBCI for a *regression* task.

We saw in previous chapters that the best way to combine evidence to form cBCIs for both the N2pc and the P300 classifiers was the LDA-cBCI approach. For this reason, it was decided to use only this form of cBCI to combine information from participants after the P300 cBCI for target detection and localisation. The outputs of the individually-tailored neural networks were combined via an LDA regressor.

As before, we tested groups of all possible sizes for each level. However, given the results obtained in the previous chapters, the cBCI was only computed for

levels that used presentation rates of 10 Hz or lower (i.e., levels 1–5).

System training The collaborative P300 classifier was trained using the LDA-cBCI approach presented in Section 3.2.5. That is, for each fold, the features for the P300 detection using the E_{20} combination were extracted for each participant and used to tailor an individual SVM classifier. The outputs of these classifiers on the training set were used to train an LDA (collaborative P₃₀₀-LDA) to assign a different weight to each member of the group¹.

A neural network regressor was individually trained for each group member in the same way as the sequential regression system (see Section 5.2.5) using the N2pc features extracted from all target trials from the training set, plus the outputs of the collaborative P₃₀₀-LDA classifier for the corresponding targets².

An LDA regressor (collaborative N2pc-LDA) was used to combine the outputs of the neural network regressors from group members. This regressor was trained using the the output of the individually tailored neural networks for all the targets from the training set and the x-coordinate of the target for each trial as the label.

System testing The main goal of this system was to determine whether the collaborative approach would help improve the (individual) sequential regression system. For this reason, the performance analysis of the collaborative sequential regression system focused only on the correlation between the actual and the predicted x-coordinates.

¹As we noted previously, if training the collaborative P₃₀₀-LDA classifier using the outputs from the P300 SVM classifier for the training set has any effect on performance, it will be to overfit to the training set, so it provides a lower boundary of performance for an unseen test set.

²Note that the N2pc features are extracted *individually* for each user, and concatenated with the *collaborative* score from the P300 detection classifier.

5.2.6 Calculation of the Information Transfer Rate

The information transfer rate (ITR) was first proposed as a metric for determining the usability of a BCI by Wolpaw *et al.* [1998], and has become one of the most commonly used metrics for this purpose [Thompson *et al.*, 2013, 2014]. In order to calculate the ITR, one assumes that the machine learning component of the BCI is a communication channel that converts the input features into an output (the output class), which can be correct or incorrect. Different properties of this channel (and, in turn, of the BCI) and its quality can then be assessed using information theory formulae.

In the formula proposed by Wolpaw *et al.* [1998], the first one in which the speed of the system is taken into account together with its accuracy to measure the performance of the BCI, the numerator of the ITR (also termed the bit rate or bits per symbol, B) is calculated according to the following formula:

$$B = \log_2 N + P \cdot \log_2 P + (1 - P) \cdot \log_2 \left(\frac{1 - P}{N - 1} \right), \quad (5.1)$$

where N is the number of classes and P is the classifier accuracy.

The ITR of the system is then typically measured in bits/minute:

$$ITR = B \cdot \frac{60}{T},$$

where T (seconds/symbol) is the time needed to output a decision.

Equation 5.1 is only valid under a series of assumptions that are not necessarily met in BCI systems [Yuan *et al.*, 2013]. Amongst these, there are two main assumptions that are not met by our systems:

-
- *All classes are equally probable.* In our system, the probability of targets is set at 10%, which means that non-target trials have a probability of 90%. Moreover, if we further divide the target class into LVF, RVF and central targets, the imbalance is even more obvious.
 - *Classification accuracy is the same for all classes.* So far we have not studied this aspect in detail to see where most misclassifications occur in our systems, but without first looking into this (as we will do using the confusion matrix in the next sections), we cannot assume that this is the case.

An alternative use to compute the bit rate of the system is through its mutual information. In general, for a system with N possible classes, the mutual information between the user intention, X , and the system output, Y , is given by

$$I(X; Y) = \sum_{i=1}^N \sum_{j=1}^N p(x_i) \cdot p(y_j|x_i) \cdot \log_2(p(y_j|x_i)) - \sum_{j=1}^N p(y_j) \cdot \log_2(p(y_j)) \quad (5.2)$$

where $p(x_i)$ is the a priory probability of class x_i and $p(y_j|x_i)$ is the probability of classifying x_i as y_j [Schlögl *et al.*, 2007]. These values can be easily computed from the confusion matrix of the system.

In this work, we decided to use this alternative way of computing the bit rate of the system to then obtain the ITR, since it does not make any of the assumptions that are present in Equation 5.1.

Table 5.1: Median AUC values across all participants for each difficulty level for the sBCI (E_{20}) and the sequential N-P system, for T vs NT classification, and Bonferroni-corrected p values (last row).

	<i>Level 1</i>	<i>Level 2</i>	<i>Level 3</i>	<i>Level 4</i>	<i>Level 5</i>	<i>Level 6</i>	<i>Level 7</i>
E_{20} sBCI	0.856	0.844	0.697	0.759	0.579	0.655	0.585
N-P system	0.860	0.843	0.683	0.768	0.584	0.646	0.579
p values	0.14	0.09	0.48	0.16	0.33	0.35	0.09

5.3 Results

5.3.1 Sequential N-P System

Table 5.1 reports the median AUC values across all participants for a each level of difficulty for the sequential N-P system together with the values obtained at each level for the equivalent E_{20} configuration in Chapter 4. It can be observed that the sequential N-P approach beats the original sBCI at levels 1, 4 and 5. A one-sided pairwise Wilcoxon signed rank test comparing mean performance for each individual across all levels within a given difficulty level revealed that *none of the differences in performance across the two conditions is statistically significant*. The p value obtained by means of this test for each level can also be seen in Table 5.1.

The mean and standard deviation of the ITR of this system are reported for each difficulty level in Table 5.2. As the table shows, higher speeds do not necessarily lead to higher ITRs. This is due to the fact that the number of errors in the system that occur as a result of increasing the speed (seen in this and previous chapters in the form of decreasing AUC values) increases much quicker than the speed of the system to output a new symbol (e.g., from 200 ms in level 1 to 66.7 ms in level 7). Thus, at least in the single-user BCI approach, there

Table 5.2: Mean and standard deviation of the ITR for the sequential N–P system, across all participants and for each difficulty level.

<i>Level</i>	<i>Mean ITR</i>	<i>SD</i>
1	43.26	22.41
2	50.19	31.47
3	29.92	21.19
4	45.85	30.43
5	19.66	10.79
6	32.67	21.92
7	19.22	13.60

is no advantage in increasing the speed of the system from the communication perspective.

Also consistently with previous results, there are marked differences between levels 2 and 3, as well as between levels 4 and 5. Again, the number of errors made in the classification of targets when changing from a unique airplane shape to several templates while keeping the same presentation rate, is increased greatly, giving rise to these differences in ITR.

5.3.2 Sequential P–N System

Table 5.3 summarises the results for the sequential P–N system. In particular, it compares the performance of this system with the single-user BCI developed in Chapter 4, in which the LVF vs RVF classifier (i.e., N2pc sBCI) was first introduced.

In order to be able to compare the new system with the N2pc sBCI alone, the median AUCs reported in Table 5.3 were calculated using only the lateral targets from the test set of each fold, as was done for the single-user BCI case in Chapter 4. The last row of the table shows the p value from one-sided pairwise

Table 5.3: Median AUC values across all participants for each difficulty level for the sBCI and the sequential P–N system, for left vs right target classification, and Bonferroni-corrected p values (last row).

	<i>Level 1</i>	<i>Level 2</i>	<i>Level 3</i>	<i>Level 4</i>	<i>Level 5</i>	<i>Level 6</i>	<i>Level 7</i>
<i>N2pc sBCI</i>	0.756	0.781	0.669	0.720	0.652	0.680	0.596
<i>P–N system</i>	0.757	0.787	0.651	0.730	0.677	0.681	0.575
<i>p values</i>	0.38	0.5	0.38	0.46	0.10	0.42	0.35

comparisons between the two systems. These show that *the small differences seen on the table are not statistically significant*. As explained in the rationale of Section 5.2.4, the performance of the N2pc classifier from the P–N system was not expected to improve with respect to the N2pc sBCI system introduced in previous chapters, so this result is not surprising.

One of the disadvantages of the sequential N–P system presented in the previous section is that information is lost throughout the system: in order to train the first part (the N2pc classifier), information regarding the position of a target in the training set is needed. However, the final output (target vs non-target) offers no indication about this aspect. The ITR of the N–P system, thus, was only computed on the final two classes — targets and non-targets.

In contrast to the N–P system, trials that go through the P–N system have a different flow depending on the decision of the first stage: the P300 classifier. If a trial is associated with a target label, it goes through the N2pc classifier, which decides whether this target is located on the left or the right hemifield. If, on the contrary, the P300 classifier decides that the trial is a non-target, that will be its final label. This system, thus, has *three possible different outcomes*, which are all included in the calculation of the confusion matrices and the ITR.

Figure 5.4 shows an example of confusion matrix, for participant 4 in level 2.

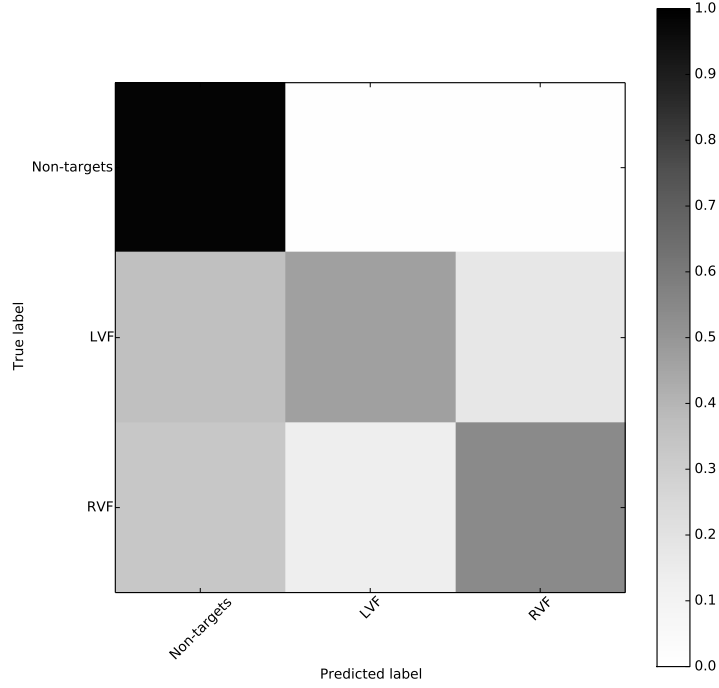


Figure 5.4: Confusion matrix from participant 3 at level 2 on the sequential P–N system.

Due to the imbalance between the classes, the values to compute the confusion matrix were normalised, so that the values across each row add up to 1. In this way, the color of the square in row i and column j shows the proportion of trials of class i that are labelled by the system as class j . Table 5.4 details the level-by-level mean and standard deviation of the ITR across all participants.

The ITRs obtained with the P–N system are, with the exception of level 2, lower than those reported on Table 5.2 for the sequential N–P system. This seems to indicate that in order to get a more sophisticated system, i.e., one that is capable of predicting the approximate location of the target, it is necessary

Table 5.4: Mean and standard deviation of the ITR for the sequential P–N system, across all participants and for each difficulty level.

<i>Level</i>	<i>Mean ITR</i>	<i>SD</i>
1	42.31	22.15
2	51.75	34.67
3	26.32	21.16
4	42.57	32.71
5	10.78	10.21
6	30.44	33.62
7	12.23	11.87

to sacrifice some of the speed. However, a look at Figure 5.4 highlights the fact that part of the problem is found in the first stage of the system¹: the classifier that decides whether a trial goes through the whole pipeline or not. Whilst around 95% of the non-target trials are correctly identified, a large proportion of targets are misclassified as non-targets in this step, thus markedly impacting the performance of the full system. Thus, we can now see that the second condition mentioned in Section 5.2.6 for using Equation 5.1 is not met either. Moreover, the P300 classifier seems to make the same number of errors, regardless of the side of the screen where a target is located.

It is interesting to notice, however, that the rate of misclassifications in the second part of the system is much lower, with only between 10–15% of targets of one type being misclassified as the other type.

5.3.3 Sequential Regression System

Table 5.5 shows, for each level of difficulty, the average Pearson’s correlation coefficient between the real and the predicted x-coordinate of the targets on the

¹Even though this confusion matrix was plotted for one participant, the results shown in it are consistent across most participants and levels.

Table 5.5: Mean and standard deviation of the correlation coefficient, ρ , and the slope (β) of the regression line fitted to the test set across all difficulty levels.

Level	$\rho (\mu \pm \text{SD})$	$\beta (\mu \pm \text{SD})$	β/ρ
1	0.19 ± 0.12	0.10 ± 0.07	0.53
2	0.26 ± 0.13	0.15 ± 0.09	0.58
3	0.08 ± 0.12	0.04 ± 0.05	0.50
4	0.19 ± 0.12	0.11 ± 0.07	0.58
5	0.13 ± 0.12	0.06 ± 0.06	0.46
6	0.16 ± 0.12	0.08 ± 0.06	0.50
7	0.09 ± 0.08	0.04 ± 0.04	0.44

test set, across all participants of each level, as well as the mean slope of a regression line fitted across the predictions.

As explained in Section 5.2.5.3, since the scaling factor between the correlation coefficient and the regression slope is the ratio between the standard deviations of the outputs and the inputs, the low slopes recorded in Table 5.5 are an indication of the smaller variance of the predictions than the variance of the x-coordinates that are given as inputs. Indeed, the ratio $\frac{\beta}{\rho}$ is maintained around 0.5 across all levels.

Despite the low average correlation coefficients reported in the table, individual users can achieve much higher correlations. The highest correlation coefficient, $\rho=0.43$, was recorded by participant 3 for level 2, who also obtained a regression slope of 0.28 (the highest slope across all participants and levels). Figure 5.5 shows the predicted vs real coordinate of all target images in the test set for this participant and level (with $\beta/\rho=0.65$).

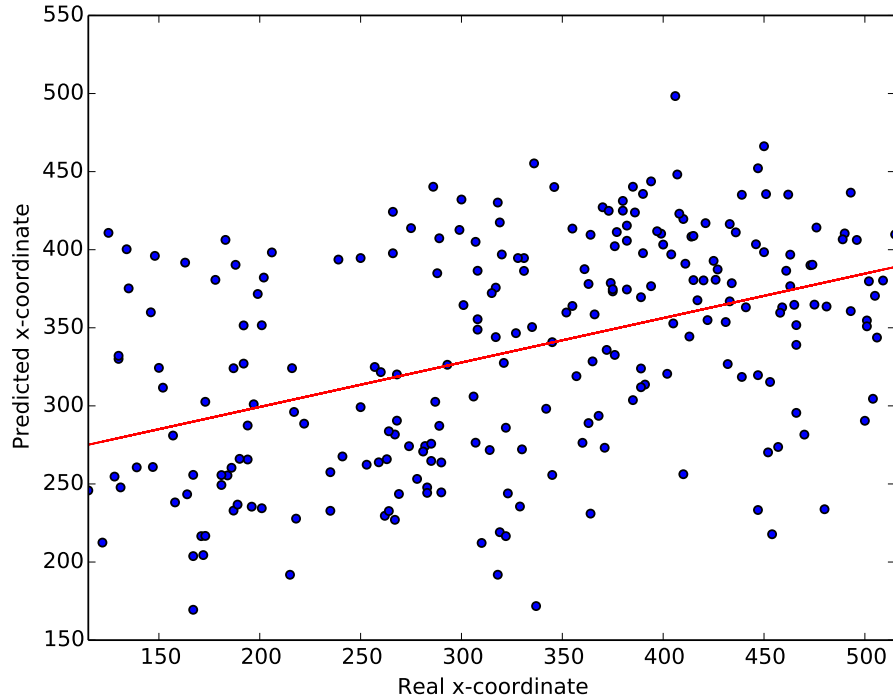


Figure 5.5: Predicted vs real x-coordinate of targets, for the best performer at this task (participant 3, in level 2).

5.3.4 Collaborative Sequential Regression System

This last section of results of the chapter reports the results obtained when using the collaborative approach at the output of the neural network of the sequential regression system.

Table 5.6 shows the mean and standard deviation of Pearson’s correlation coefficient for levels 1–5 and all possible group sizes. Similarly, the regression slope is reported in Table 5.7. In general, both the correlations and the regression slopes increase with group sizes. Moreover, the highest mean values are recorded for level 2, and then decrease with increasing levels of difficulty, in the same way as the shape of the N2pc-based cBCI behaved.

Table 5.6: Mean and standard deviation of the correlation coefficient between actual and predicted x-coordinate of targets for different group sizes.

Size	Level 1	Level 2	Level 3	Level 4	Level 5
2	0.40 ± 0.09	0.44 ± 0.13	0.25 ± 0.11	0.33 ± 0.14	0.22 ± 0.11
3	0.46 ± 0.08	0.51 ± 0.10	0.29 ± 0.10	0.40 ± 0.11	0.27 ± 0.09
4	0.51 ± 0.07	0.56 ± 0.08	0.32 ± 0.09	0.44 ± 0.09	0.30 ± 0.09
5	0.54 ± 0.06	0.59 ± 0.07	0.34 ± 0.08	0.47 ± 0.07	0.33 ± 0.08
6	0.57 ± 0.05	0.62 ± 0.05	0.36 ± 0.07	0.50 ± 0.06	0.35 ± 0.06
7	0.59 ± 0.04	0.64 ± 0.04	0.37 ± 0.07	0.51 ± 0.05	0.37 ± 0.05
8	0.60 ± 0.04	0.66 ± 0.04	0.39 ± 0.06	0.53 ± 0.04	0.39 ± 0.04
9	0.62 ± 0.03	0.68 ± 0.03	0.39 ± 0.04	0.54 ± 0.02	0.39 ± 0.00

Table 5.7: Mean and standard deviation of the slope of the regression line fitted to the test set for different group sizes and difficulty levels.

Size	Level 1	Level 2	Level 3	Level 4	Level 5
2	0.37 ± 0.09	0.42 ± 0.14	0.20 ± 0.09	0.31 ± 0.13	0.17 ± 0.09
3	0.39 ± 0.07	0.45 ± 0.10	0.20 ± 0.08	0.33 ± 0.10	0.18 ± 0.06
4	0.40 ± 0.06	0.46 ± 0.08	0.21 ± 0.06	0.34 ± 0.08	0.19 ± 0.05
5	0.40 ± 0.05	0.47 ± 0.07	0.21 ± 0.05	0.35 ± 0.07	0.19 ± 0.04
6	0.41 ± 0.04	0.48 ± 0.05	0.22 ± 0.04	0.35 ± 0.06	0.19 ± 0.03
7	0.41 ± 0.03	0.48 ± 0.05	0.22 ± 0.04	0.36 ± 0.05	0.19 ± 0.03
8	0.41 ± 0.03	0.49 ± 0.04	0.22 ± 0.04	0.36 ± 0.04	0.20 ± 0.02
9	0.41 ± 0.03	0.49 ± 0.03	0.21 ± 0.03	0.35 ± 0.02	0.18 ± 0.00

As we saw before, these two parameters are closely related through the variance of the system outputs. Table 5.8 reports the ratios between the mean regression slopes and the mean correlation coefficients for each level and group size. The highest ratios are obtained, once again, for level 2. Interestingly, the ratios now *decrease* with increasing group sizes, revealing that the standard deviation of the outputs decreases (with respect to the *constant* standard deviation of the inputs).

Table 5.8: Ratio between the mean regression slope and mean correlation coefficients between the predicted and the actual x-coordinates of targets.

Size	Level 1	Level 2	Level 3	Level 4	Level 5
2	0.93	0.95	0.80	0.94	0.77
3	0.85	0.88	0.69	0.83	0.67
4	0.78	0.82	0.66	0.77	0.63
5	0.74	0.80	0.61	0.74	0.58
6	0.72	0.77	0.61	0.7	0.54
7	0.69	0.75	0.59	0.71	0.51
8	0.68	0.74	0.56	0.68	0.51
9	0.66	0.72	0.54	0.65	0.46

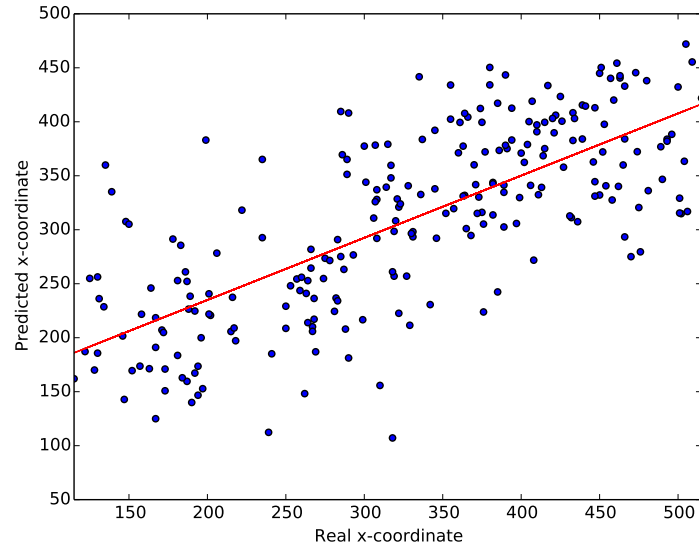


Figure 5.6: Predicted vs real x-coordinate of targets, for a group of size 7 at level 2.

Figure 5.6 shows the predicted vs the real coordinate of all target images in the test set of level 2 for one group of size 7 with a correlation coefficient of 0.72 (the highest obtained all throughout the results) and a slope of 0.58.

5.4 Discussion

This chapter looked at the possibility of exploiting the P300 and N2pc ERPs together in a BCI which is capable of jointly detecting *and* locating targets in rapid streams of pictures.

In relation to the N–P system, we have seen that adding a shift of attention classifier (i.e., the N2pc) before trying to detect a target does not improve detection results. This might be due to the fact that the N2pc classifier will give an output score that is correlated to the position of the target and, because of this, it will vary proportionally with the x-coordinate of the airplane in the picture. That is, the range of outputs of the N2pc classifier (near zero for both non-targets and central targets, and ± 1 for lateral targets, depending on which side of the image they are located) does not allow the P300 classifier (an ensemble of linear SVMs) to learn a rule regarding the presence or absence of a target. Thus, this extra feature that the P300 classifier receives, with respect to the E_{20} configuration used in previous chapters, seems to be disregarded by the system.

The sequential P–N system attempted the opposite: it first detected whether there was a target in the image before trying to locate it within it. Again, results showed that this did not improve the systems presented in the previous chapter to locate the target in the image, a result that was, somehow, expected.

It was hoped that changing from a 2-class problem (i.e., T vs NT in both the N–P system and the original target detector from Chapter 3) to a 3-class problem (i.e., NT vs LVF vs RVF in the P–N system) would increase the information transfer rate of the system. However, the addition of the target classifier before deciding the location of the target did not achieve this effect due to the large

number of misclassified targets of the P300 detector, even at low presentation rates.

Thanks to the use of the confusion matrices, it was observed that errors in the target detection system occur in targets regardless of their location (i.e., left or right side of the image). In contrast, errors within the N2pc classifier are less frequent. This suggests that the target detector is the system that can benefit the most from improvements (e.g., by pooling information from multiple participants to create a collaborative BCI).

In the sequential regression system, the N2pc classifier was substituted with a neural network regressor that was capable of increasing the correlation coefficient between the predicted and the actual coordinate of the target with respect to that obtained with the PSO-based linear regressor that was introduced in Chapter 4.

The sequential regression system was further improved by creating a collaborative regressor. Bigger groups exhibited higher correlations and regression slopes between the predicted and the real coordinate of the object of interest (an airplane) within an image. By creating groups of users, the variance in the output of the system (i.e., the predicted coordinate) was reduced, which is possibly the reason for the higher correlation coefficients and regression slopes. Due to this, it is now possible to predict the actual position of targets in images through the collaborative version of the sequential regression system.

5.5 Conclusions

This chapter attempted the task of combining the target detection and localisation systems developed in previous chapters.

The sequential P–N and N–P systems did not improve the performance of the single-user BCI systems presented in the previous chapters. However, we obtained better localisation of the targets by using a sequential regressor system based on a neural network than we did using a PSO. These improvements were further increased in the collaborative approach of the sequential regression system. Even though the localisation system (based on a PSO) presented in the previous chapter was capable of highlighting the area where the target was located in the picture, it is thanks to the collaborative sequential regressor system that we can now predict the x-coordinate of the target.

Considering that target detection seems to be the bottle neck for the sequential P–N systems, and that the collaborative systems presented in Chapter 3 were significantly better at target detection than the single-user BCIs, we would expect that the amount of misclassifications from the P300 detection system would be much lower in the collaborative sequential regression system, thus improving the ITR of the system overall. However, we did not explore here to what extent this is true, and it remains a task for future work.

Moreover, we must note that the systems presented here use data that has been re-used in cross-validation loops multiple times at this point, so it could be that our systems are evolving solutions that overfit our dataset. Thus, the viability of these systems should be further tested on new datasets.

One of the main aims of this PhD thesis is to see where the improvements of performance that are obtained by merging information from multiple users is coming from. In the next chapter, we will study the effects of selecting group members based on their individual performance, and develop a theoretical model that partly explains the improvements that are obtained in collaborative BCIs.

Chapter 6

Improving cBCI Performance with Participant Selection

This chapter introduces a method for creating groups based on performance similarity between the group members. It also explores some possible reasons for the improvements obtained by the cBCIs over the sBCI approach reported in previous chapters and presents a theoretical model that can partly explain some of these improvements.

6.1 Introduction

We have seen that there are advantages in forming groups in collaborative BCIs for detecting and locating targets in images. This chapter explores the effects of *selecting* the participants which form the groups. This will be done on the basis of performance similarity between group members. The reason for testing this idea instead of the more obvious notion that groups with higher-AUC members

perform better than groups with lower-AUC members is the research reported by Bahrami *et al.* [2010], where the authors showed that pairs of participants perform better than individual observers, provided that they had similar visual sensitivities and were able to communicate. This chapter postulates that the better AUCs obtained by some members are largely attributed to the fact that their perceptual system had a higher sensitivity (whether this is at the preattentive, attentive or cognitive level remains as a task for future work).

We saw in previous chapters that performance varies widely across participants for a given level of difficulty, and that performance improves, on average, when using the collaborative BCI approach. However, the reasons behind these improvements have not been studied in depth.

This chapter will present a group member selection mechanism that takes into account the AUC of each individual and its similarity with that of the other candidates to be group members in order to decide whether he/she should be included in the group. It will also compare the effects of applying this selection method with two systems: one, that behaves as the average participant of the group; and another that behaves as the best performer of the group.

Finally, a theoretical model is introduced that can largely explain the improvements obtained when using cBCIs with respect to the sBCI systems when combining evidence from multiple viewers, both when the groups are formed freely (as was done in the previous chapters) and when selecting them based on AUC similarity as we do in this chapter.

6.2 Methods

6.2.1 Group Member Selection

In relation to the selection of group members, the method presented here forms group according to the similarity in performance (i.e., AUC) of individual participants, using different levels (or thresholds) of similarity. More specifically, participants are allowed to form a group if the range of AUCs across the participants for a given classification task, a value that we term *dissimilarity index*, was below a threshold δ . That is, if participants i and j were the ones with the highest and lowest AUC values, the group could be formed if

$$|AUC_i^f - AUC_j^f| \times 100\% \leq \delta,$$

where AUC_x^f represents the AUC value for participant x (with $x = 1, \dots, 11$) at level of difficulty f (with $f = 1, \dots, 7$). Groups were created by setting the threshold δ at 5, 10, 15, 20 and 25% and considered only the cBCIs obtained from groups of subjects for which the dissimilarity index was below the threshold. For comparison, the situation where no group selection was performed (i.e., $\delta = 100\%$) and all possible groups of a given size are included is also considered.

Of course, this selection process reduces the number of groups that can be included in the analysis. However, given that collaborative BCIs are conceived with the aim of augmenting human capabilities, it is reasonable to select participants based on their individual performance when forming groups.

6.2.2 Data

The group member selection method described above was applied to two scenarios: the target detection system introduced in Chapter 3 (Section 6.3.1) and the target localisation system studied in Chapter 4 (Section 6.3.2).

Thus, this chapter reports the results for each level of difficulty and group size for each of these methods (computed as the median AUC obtained across all the groups included for each value of the dissimilarity index). Even though this chapter describes the main results, with examples for each system, due to the large number of tables needed to report such results, most of these have been placed in Appendix C.

In order to avoid forming groups using data that is later used to measure the amount of the improvement with respect to the single-user case, in this chapter the data were divided into a training and a test set as follows:

- The training set contained 75% of the data. This set was used to perform stratified 10-fold cross-validation in order to measure the cross-validation AUC and to select the parameters of the classifiers. The average AUC on cross-validation from each participant and level was used to determine his/her similarity with other candidates to form a group.
- The test set contained the remaining 25% of the data, and was only used to compare the effects of forming groups on *unseen* data, as explained below. For this, the classifiers were trained using all the data from the training set.

The train-test split was performed so that the ratio of trials of each class was approximately the same in the train and the test sets. Training and test datasets

were created individually for the target detection and the target localisation systems.

In addition to the actual improvements of performance achieved by the group member selection procedure, this chapter also shows the effects of this method by considering the fraction of groups (out of the total number of groups of each size that can be formed) that are included for each group size and value of the dissimilarity index.

6.2.3 Comparisons with Other Systems

For a deeper analysis of the degree to which a cBCI provides improvements over individual sBCI performance, the results of applying group selection in the target detection and target localisation systems were compared to two reference systems for making joint decisions: (1) one unintelligent system that chooses, at random, to follow the classification decisions provided by the sBCI associated with one member of a group, and (2) one, more intelligent, system that always chooses the decisions of the better performing individual in a group. Obviously, the AUCs obtained in these two systems would be, for a group of size r , the $\text{avg}(AUC_1, AUC_2, \dots, AUC_r)$ for the former, and $\text{max}(AUC_1, AUC_2, \dots, AUC_r)$ for the latter, where AUC_i represents the AUC of the sBCI adapted to participant $i \in [1, 11]$.

The reference systems use the *test* AUC from the participants that formed the groups and compare them to the AUC obtained by the group in the *test* set.

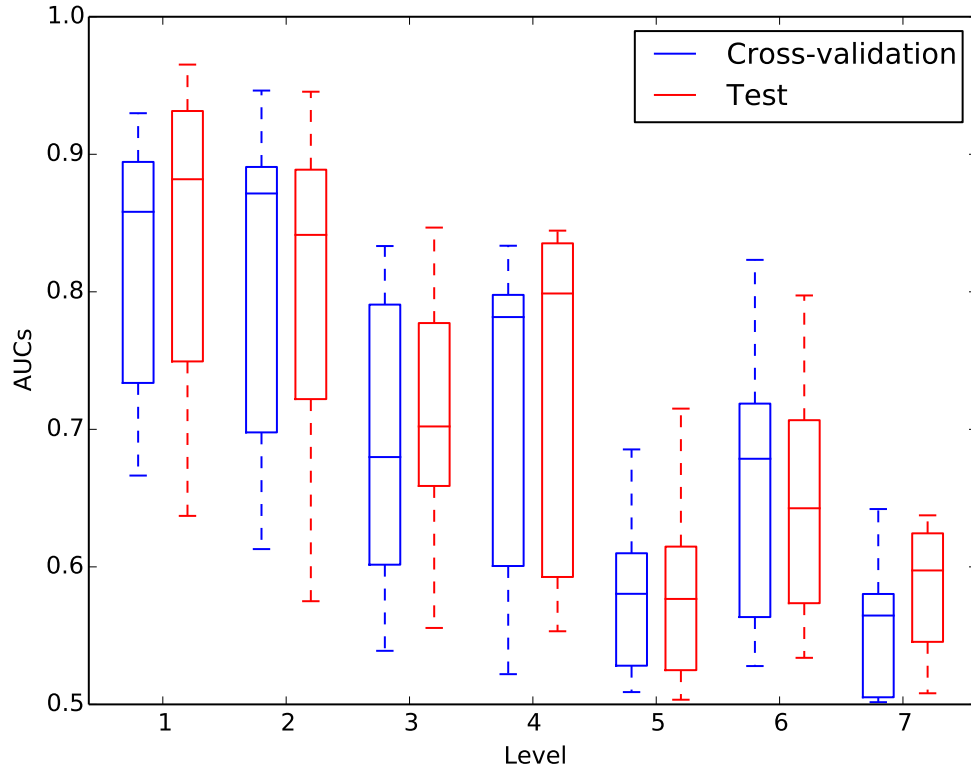


Figure 6.1: Distributions of mean AUCs for the cross-validation and test folds for the single-user target detection BCI across all levels of the experiment.

6.3 Results

6.3.1 Participant Selection in the Target Detection System

Before looking at the effects of forming groups, it is worth noting that, as shown in Figure 6.1, the AUCs obtained for each level are not significantly different between the cross-validation and test sets. This was assessed via a level-by-level one-sided paired Wilcoxon signed rank test, in which all p values were higher than 5% *before* applying Bonferroni correction.

Table 6.1 reports the median AUC values for the SC-cBCI, MC-cBCI and

Table 6.1: Median AUC values for the three types of cBCIs for target vs non-target classification for difficulty level 2, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	0.93	0.93	0.93	0.93	—	—	—	—	—
	10%	0.93	0.95	0.95	0.96	0.96	—	—	—	—
	15%	0.92	0.94	0.95	0.95	0.95	—	—	—	—
	20%	0.92	0.94	0.95	0.96	0.96	0.97	—	—	—
	25%	0.89	0.92	0.93	0.93	0.94	0.94	0.95	—	—
	100%	0.88	0.91	0.92	0.93	0.94	0.95	0.95	0.95	0.96
MC-cBCI	5%	0.94	0.95	0.97	0.97	—	—	—	—	—
	10%	0.94	0.96	0.97	0.98	0.98	—	—	—	—
	15%	0.93	0.95	0.97	0.97	0.97	—	—	—	—
	20%	0.92	0.95	0.97	0.97	0.98	0.98	—	—	—
	25%	0.91	0.94	0.96	0.96	0.97	0.97	0.98	—	—
	100%	0.90	0.93	0.94	0.96	0.96	0.97	0.97	0.97	0.97
LDA-cBCI	5%	0.93	0.95	0.96	0.97	—	—	—	—	—
	10%	0.93	0.96	0.97	0.97	0.97	—	—	—	—
	15%	0.92	0.95	0.96	0.97	0.97	—	—	—	—
	20%	0.92	0.95	0.97	0.97	0.97	0.98	—	—	—
	25%	0.91	0.94	0.96	0.96	0.97	0.97	0.97	—	—
	100%	0.91	0.94	0.95	0.96	0.97	0.97	0.97	0.98	0.98

LDA-cBCI methods of combining evidence to form collaborative BCIs, for different group sizes and as a function of δ , for level 2. Empty values of the table represent cases where the value of the dissimilarity index did not allow any groups to be made.

Comparing this table with the median single-user BCI AUC of 0.86 reported on Figure 3.6 and Figure 6.1, it can be seen that AUCs are markedly higher (by up to almost 13%) than for single-user BCIs, for all types of collaborative BCIs. As shown numerically in this table, performance of cBCIs generally decreases with

Table 6.2: Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 2.

δ	Group size								
	2	3	4	5	6	7	8	9	10
5%	26%	8%	2%	0%	0%	0%	0%	0%	0%
10%	44%	18%	7%	2%	0%	0%	0%	0%	0%
15%	55%	26%	11%	4%	0%	0%	0%	0%	0%
20%	62%	33%	17%	8%	3%	0%	0%	0%	0%
25%	82%	61%	43%	28%	16%	7%	2%	0%	0%

increasing dissimilarity indices, but, even when no pair selection is performed (i.e., $\delta=100\%$), as we saw in Chapter 3, cBCIs are better than the corresponding sBCI.

With N participants, we can form up to $\binom{N}{r}$ distinct groups of size r . Table 6.2 reports the effects that different values of the dissimilarity-index threshold (see Section 6.2.1) have on the fraction of groups that can be accepted for level 2. Obviously, all groups are accepted for $\delta=100\%$, so this case is not reported on the table.

Section C.1.1 contains the results (median AUC values for all group sizes and cBCI methods, together with the fractions of groups accepted) for the remaining levels of the target detection system.

6.3.1.1 Comparison with $\text{avg}(AUC_1, AUC_2, \dots, AUC_r)$

Table 6.3 reports the median gains in AUC over the average performance of the individuals in each group for each group size r , separately for the three types of collaborative BCIs — single-classifier cBCIs (SC-cBCI), multiple-classifier cB-

Table 6.3: Median improvements over the average participant in the group when using collaborative BCIs for target detection at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group).

Method	δ	Group size				
		2	3	4	5	6
SC-cBCI	5%	+4.9%	+6.4%	+7.7%	+8.7%	–
	10%	+4.4%	+6.3%	+7.2%	+7.8%	+8.9%
	15%	+5.2%	+7.1%	+7.9%	+8.7%	+10.2%
	20%	+5.4%	+7.6%	+9.3%	+10.2%	+10.3%
	25%	+5.7%	+9.4%	+11.0%	+12.9%	+14.0%
	100%	+6.6%	+11.0%	+13.2%	+14.7%	+16.1%
MC-cBCI	5%	+6.2%	+9.2%	+10.1%	+10.5%	–
	10%	+6.2%	+8.7%	+9.1%	+9.5%	+9.7%
	15%	+7.4%	+9.5%	+10.1%	+10.5%	+11.7%
	20%	+7.6%	+10.1%	+11.0%	+12.0%	+12.5%
	25%	+8.0%	+11.6%	+13.7%	+15.2%	+16.4%
	100%	+9.4%	+13.4%	+16.1%	+17.5%	+18.6%
LDA-cBCI	5%	+6.3%	+9.3%	+10.5%	+10.6%	–
	10%	+6.5%	+8.8%	+9.0%	+9.7%	+9.8%
	15%	+7.6%	+9.8%	+10.5%	+10.6%	+11.8%
	20%	+8.2%	+10.6%	+11.2%	+12.3%	+12.8%
	25%	+9.3%	+13.0%	+14.7%	+16.3%	+17.4%
	100%	+10.7%	+15.0%	+17.7%	+18.8%	+20.1%

CIIs (MC-cBCI) and LDA-cBCIs — for different values of the dissimilarity-index threshold δ .

Unsurprisingly, all values in the table are positive, indicating that, irrespective of the value of δ , cBCIs outperform the unintelligent system that randomly picks the responses of one individual in a group. These improvements are highly statistically significant in almost every case. These results are generalised across all levels for the target detection task. Tables reporting the results of comparing the achieved improvements with the average AUC of the group members, across the remaining levels and for all group sizes can be found in Section C.1.2.

Table 6.4: Median improvements over the best participant in the group when using collaborative BCIs for target detection at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size				
		2	3	4	5	6
SC-cBCI	5%	+2.0%	+2.1%	+2.7%	+3.0%	–
	10%	+0.3%	+0.9%	+1.3%	+1.4%	+2.1%
	15%	+0.6%	+0.9%	+1.3%	+1.5%	+2.5%
	20%	<i>+0.3%</i>	+0.8%	+0.9%	+1.0%	+0.9%
	25%	-0.0%	+0.3%	+0.8%	+1.2%	+1.4%
	100%	-0.9%	-1.2%	-0.6%	-0.4%	-0.1%
MC-cBCI	5%	+3.3%	+4.1%	+4.4%	+4.8%	–
	10%	+2.0%	+2.8%	+2.6%	+2.8%	+3.0%
	15%	+2.2%	+3.0%	+3.1%	+4.1%	+3.9%
	20%	+2.0%	+2.5%	+2.7%	+2.6%	+2.6%
	25%	+1.4%	+2.3%	+2.8%	+2.9%	+3.1%
	100%	+0.7%	+1.0%	+1.4%	+1.7%	+1.6%
LDA-cBCI	5%	+3.4%	+4.1%	+4.7%	+4.9%	–
	10%	+2.1%	+3.0%	+2.7%	+2.8%	+3.0%
	15%	+2.2%	+3.2%	+3.5%	+4.2%	+4.0%
	20%	+2.2%	+3.0%	+2.8%	+2.7%	+2.7%
	25%	+2.0%	+3.0%	+3.5%	+3.7%	+4.0%
	100%	+1.6%	+2.1%	+2.4%	+2.5%	+2.5%

6.3.1.2 Comparison with $\max(AUC_1, AUC_2, \dots, AUC_r)$

We will now study the much more challenging scenario represented by the second reference system. In this case, the performance of a cBCI is compared with that of the best participant of the group. Table 6.4 reports, for each method, the median gains in performance over the best participant of each group, for each group size r and different values of δ .

Again, we see a predominance of positive values (98% of the entries) in the table, indicating that cBCIs tend to produce better AUCs than the “intelligent” reference system too. In this case, however, we see a dependency of performance on δ : for a fixed group size, the bigger gains are obtained for the smaller values

of the dissimilarity index.

Due to space limitations, Table 6.4 shows only a subset of group sizes for level 2. The full table with the results of comparing the improvements of performance of the group selection method vs the best participant of the group, for this level and the results for the remaining difficulty levels can be found in Section C.1.3.

The results reported here are generalised across levels 1–4. However, for higher difficulty levels (i.e., levels 5–7), at high values of δ , the drops of performance (when compared with the best participant) are higher and more frequent. For example, they are present in the LDA-cBCI method in difficulty levels 5 (for small group sizes and $\delta \geq 20\%$) and 7 (for groups of size 2 and $\delta \geq 10\%$).

The dependency of performance on δ is further illustrated graphically in Figure 6.2, which shows an interpolation (via the “natural neighbour” algorithm) of the improvements obtained by the MC-cBCI method over the AUC of the best performer of each group (of size 3), including data from all difficulty levels. By plotting the improvement *versus* the mean and the standard deviation of the AUCs in each group, the low values of the dissimilarity index δ are found at the bottom of the plot, while groups formed by very different members (i.e., high values of δ) are located at the top of the figure.

This figure shows quite clearly that higher improvements are concentrated towards the bottom of the plot, thus indicating that the biggest improvements are associated with higher similarity between the participants’ AUCs, as was shown numerically above.

The reduction in cBCI performance associated with higher values of the dissimilarity index seems reasonable, considering that when participants with high

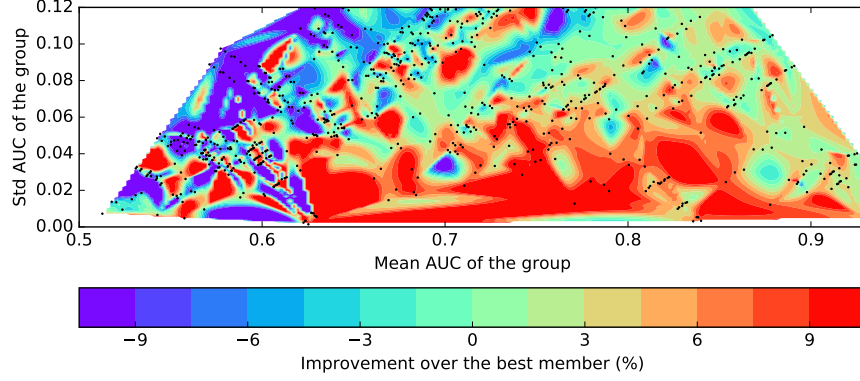


Figure 6.2: Surface interpolation of the AUC improvements obtained when using MC-cBCI for groups of size 3 over the AUC of the best member of the group.

AUCs are grouped with low scorers (thus leading to a high dissimilarity index), the limited information provided by the good individuals or group of them with respect to the bad ones is unlikely to translate into an improvement in the performance of the best participant of the group.

6.3.2 Participant Selection for Target Localisation

We will now look at the results obtained when applying the dissimilarity-index threshold δ to the collaborative BCIs for target localisation which were introduced in Chapter 4.

Following the same order as in Section 6.3.1, let us first look at the distributions of cross-validation and test AUCs that are obtained across all users for each level. These are reported in Figure 6.3. Again, it can be seen that there are no statistically significant differences between performance in the cross-validation and the test sets, except for those obtained at level 5 (for a presentation rate of 10 Hz and multiple target templates), where the performance is significantly higher (but

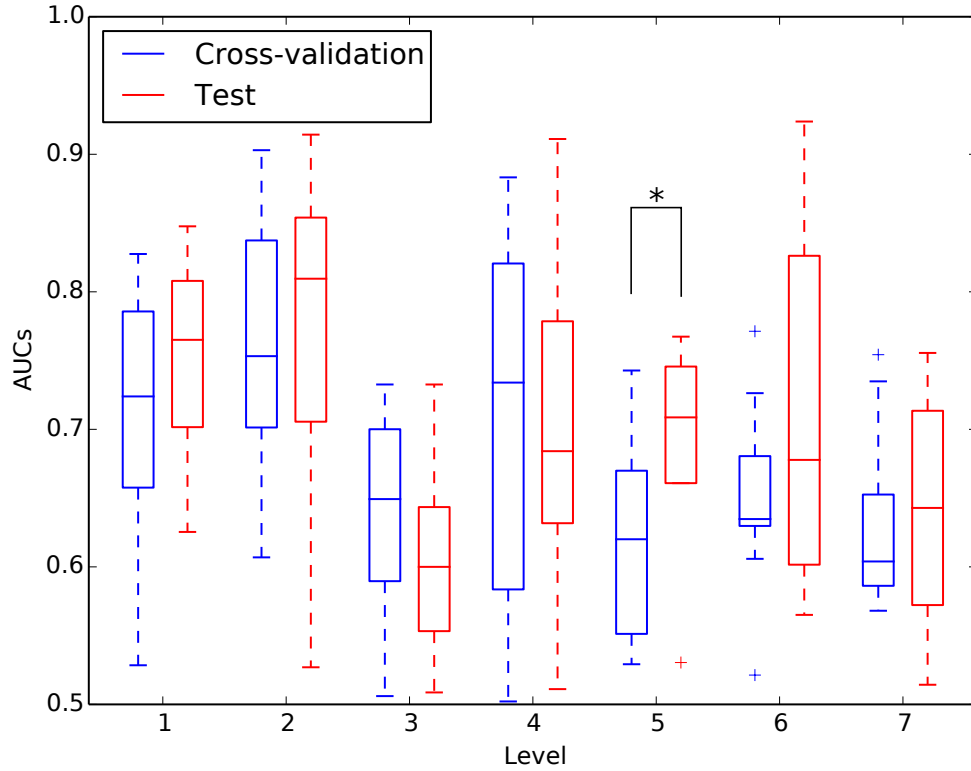


Figure 6.3: Distributions of mean AUCs for the cross-validation and test folds for the single-user target localisation BCI across all levels of the experiment.

not after Bonferroni correction) in the test set than during cross-validation.

We can now look at the results that are obtained for each cBCI method, across all values of δ and group sizes, for level 2. These can be found in the form of median AUCs on Table 6.5. The percentages of groups (out of all the possible combinations) that are considered for each group size and δ are reported in Table 6.6.

Table 6.5 shows median AUCs much higher (up to almost 26%) than the one from the sBCI used for target localisation in Chapter 4, which was 0.78 for difficulty level 2. The only exception is found for pairs of users in the SC-cBCI method. We elaborate on this effect and one possible cause for it in Sec-

Table 6.5: Median AUC values for the three types of cBCIs for left vs right classification of targets for difficulty level 2, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	0.82	0.79	—	—	—	—	—	—	—
	10%	0.82	0.82	0.88	0.92	—	—	—	—	—
	15%	0.79	0.78	0.82	0.83	0.83	—	—	—	—
	20%	0.78	0.79	0.83	0.85	0.90	0.92	—	—	—
	25%	0.76	0.77	0.79	0.82	0.83	0.83	0.84	0.86	—
	100%	0.76	0.77	0.78	0.79	0.81	0.82	0.85	0.86	0.84
MC-cBCI	5%	0.92	0.95	—	—	—	—	—	—	—
	10%	0.91	0.96	0.99	0.99	—	—	—	—	—
	15%	0.90	0.94	0.96	0.97	0.92	—	—	—	—
	20%	0.90	0.93	0.96	0.97	0.98	0.99	—	—	—
	25%	0.90	0.93	0.96	0.97	0.97	0.98	0.98	0.98	—
	100%	0.90	0.93	0.96	0.97	0.98	0.99	0.99	0.99	0.99
LDA-cBCI	5%	0.92	0.95	—	—	—	—	—	—	—
	10%	0.92	0.96	0.99	0.99	—	—	—	—	—
	15%	0.90	0.94	0.96	0.97	0.92	—	—	—	—
	20%	0.91	0.95	0.96	0.98	0.98	0.99	—	—	—
	25%	0.90	0.93	0.96	0.97	0.98	0.99	0.99	0.99	—
	100%	0.90	0.93	0.96	0.98	0.98	0.99	0.99	1.00	1.00

Table 6.6: Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 2.

δ	Group size								
	2	3	4	5	6	7	8	9	10
5%	17%	1%	0%	0%	0%	0%	0%	0%	0%
10%	48%	16%	3%	0%	0%	0%	0%	0%	0%
15%	71%	40%	17%	5%	0%	0%	0%	0%	0%
20%	84%	62%	40%	22%	9%	2%	0%	0%	0%
25%	95%	87%	76%	63%	50%	35%	22%	10%	0%

tion 6.3.2.1. As we saw in previous chapters, bigger group sizes improve the performance of the cBCI further.

Finally, and also consistent with results from the previous sections, performance of cBCIs generally decreases with increasing dissimilarity indices, but, even when no pair selection is performed, cBCIs are better than the corresponding sBCI.

6.3.2.1 Effects of Individuals on Group Performance

We saw in Chapter 4 that, even though large groups formed with the SC-cBCI method exhibited the same behaviour as the other types of cBCI, there was a marked drop in performance in the SC-cBCI method for groups of size 2 (see Figure 4.13).

In order to understand this drop, it is instructive to look at the low percentages of accepted groups for each group size reported on Table 6.6. These are indicators of the large differences in performance that are found in the cohort of participants for the target localisation task.

Even though both the P300 and the N2pc systems show relatively large interquartile ranges, the distributions of the target detection system are, in general, skewed toward high AUC values (see Figure 3.6), whereas the target localiser shows more symmetric distributions (see Figure 4.8). As an example, Figure 6.4 shows the probability density functions (pdfs) of AUCs for the target detection and target localisation systems at level 2.

It is interesting to look at the distributions shown here. Under the assumption that the AUC of a participant is, indeed, an indicator of their visual system sensitivity, the differences between the two distributions of AUCs are signalling

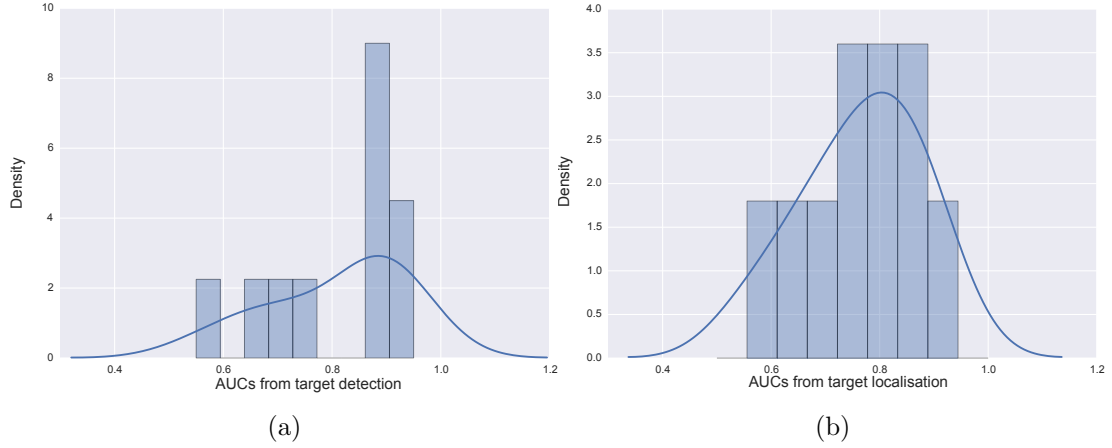


Figure 6.4: Probability density functions and histograms for the AUCs of the target (a) detection and (b) localisation systems, at difficulty level 2.

a very interesting effect. Let us first look at the distribution of AUCs for the N2pc-based BCI, shown in Figure 6.4(b), which shows a symmetrical distribution, that appears to be approximately Gaussian (i.e., the pdf superimposed over the histogram).

The N2pc ERP is an indicator of a shift of attention towards an event of interest that occurs before a person foveates towards the event. Contrary to the P300, the appearance of the N2pc does not require intentional control (i.e., the P300 is a more “active” task, enhanced by the task of counting targets in our case). For this reason, a priori, the classifier trained to distinguish between LVF and RVF targets based on this neural signature would have a homogeneous response (i.e., the AUCs will mostly be symmetrically distributed around the mean). This is what the plot indicates, with deviations from the mean being representative of noise both in the EEG and the learning rule from the classifier.

On the other hand, when the neural response depends on the participant being attentive and engaged in the task, as is the case of the P300, there are

other factors that affect the classifiers' behaviour. The pdf of the AUCs for the target detection system, shown in Figure 6.4(a), is much more skewed, reflecting actual differences in the participants' attitudes toward the task (i.e., their level of engagement, attentional blink, and his/her visual system's acuity), and not only in the classifier.

Figure 6.5 shows the pdfs of $|AUC_i - AUC_j|$ for difficulty level 2 and for groups of size 2, for both systems. There are two reasons why these plots are not just the result of convolving the distributions in Figure 6.4: (1) when computing δ the sampling is done without replacement, and (2) we are only keeping the absolute value of the difference, and not the differences themselves.

For the target localisation system (shown in Figure 6.5(b)), the peak of the distribution is narrow and centered around $\delta=0.1$. In contrast, the target detection system (in Figure 6.5(a)) shows a more uniform distribution. Hence, if one selects two participants at random to form a pair, they are more likely to have a higher dissimilarity index in the target detection system. Hence, the percentages of groups accepted are lower for this system than they are for the target localisation system for small values of δ . For example, for difficulty level 2 (see Table 6.6), 71% of all possible pairs are accepted for the target localisation system for $\delta \leq 15\%$, whereas just 55% (see Table 6.2) are included when forming pairs in the target detection system.

If we now focus on low values of δ , it can be inferred from these pairs of plots that, due to the symmetry of the distribution of AUCs, if one picks a random pair of performers, it is as likely to be formed by good performers as it is to be formed by bad ones in the N2pc-based system. However, this is not true in the case of the target detection task, where the distribution of AUCs shows a clear skew

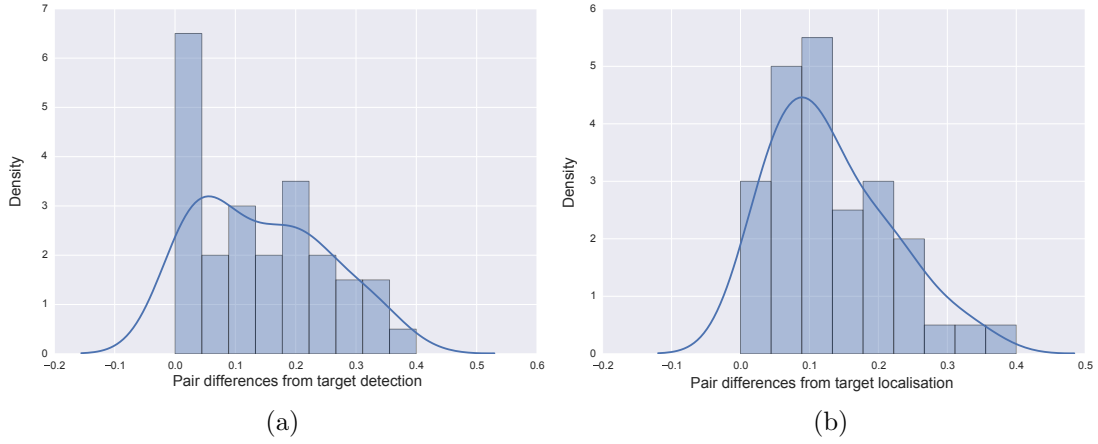


Figure 6.5: Probability density functions and histograms for the differences in pairs of AUCs (i.e., $|AUC_i - AUC_j|$) from the target (a) detection and (b) localisation systems, at difficulty level 2.

towards high values, hence favouring the occurrence of pairs of good performers. Figure 6.6 shows the pdf's and histograms of the average AUCs of pairs of users for which $\delta < 0.05$. The median values of these distributions are 0.90 for target detection and 0.79 for the target localisation system.

These results suggest an explanation for the drop in group performance observed in pairs in the SC-cBCI approach. Regardless of the task, even though pairs of good performers will (on average) show an increase of the group AUC with respect to the mean AUC of their individually tailored sBCIs, when the EEG signals from two bad performers are averaged (which, as we just saw, is more likely to happen in the N2pc classification system) to form a group in the SC-cBCI approach, there is a higher chance of them detecting/missing different targets. Hence, the SC-cBCI system will, on average, show a decrease in performance for small group sizes, as was observed in Chapter 4, due to a decrease (or a lack of increase) of the signal-to-noise ratio.

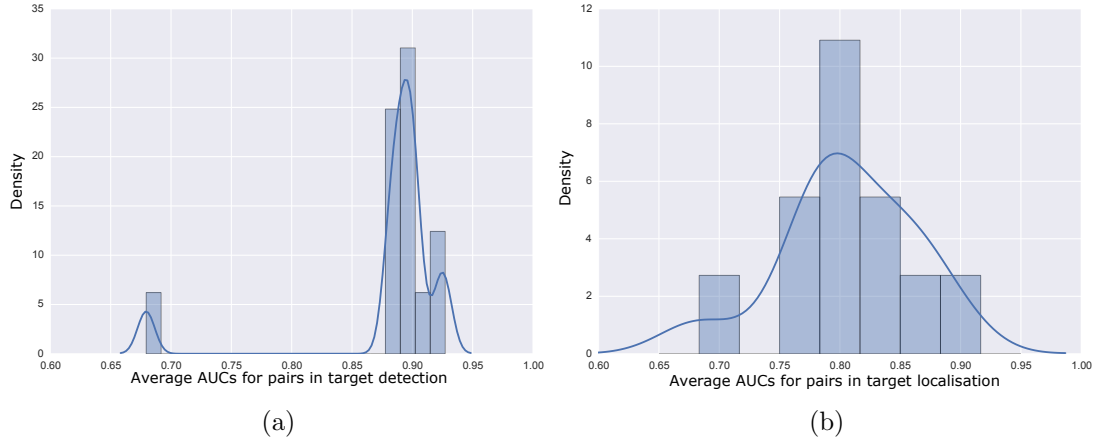


Figure 6.6: Probability density functions and histograms for the average AUCs of pairs of users (i.e., $\text{avg}(AUC_i, AUC_j)$) from the target (a) detection and (b) localisation systems, at difficulty level 2, when only pairs of users with $\delta < 0.05$ are considered.

6.3.2.2 Comparison with $\text{avg}(AUC_1, AUC_2, \dots, AUC_r)$

Following the structure from the previous section, Table 6.7 reports the median gains in AUC over the average performance of the individuals in each group for each group size r , separately for the three types of collaborative BCIs — single-classifier cBCIs (SC-cBCI), multiple-classifier cBCIs (MC-cBCI) and LDA-cBCIs — for different values of the dissimilarity-index threshold δ for the target localisation system.

With the exception of small group sizes for the SC-cBCI method, all values in the table are positive, indicating that, irrespective of the value of δ , cBCIs outperform the unintelligent system that randomly picks the responses of an individual in a group. These improvements are highly statistically significant (highlighted in bold face) in almost every case, specially for the MC-cBCI and LDA-cBCI approaches. These results are generalised across all levels, and can be

Table 6.7: Median improvements over the average participant in the group when using collaborative BCIs for target average at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test group AUC vs average

Method	δ	Group size						
		2	3	4	5	6	7	8
SC-cBCI	5%	-8.9%	-8.5%	—	—	—	—	—
	10%	-5.2%	-3.3%	-1.7%	+4.1%	—	—	—
	15%	-5.2%	-2.7%	-0.1%	+0.1%	-2.4%	—	—
	20%	-3.5%	-1.4%	+0.8%	+3.5%	+5.7%	+7.6%	—
	25%	-2.8%	-0.3%	+2.5%	+5.9%	+6.9%	+9.2%	+10.3%
	100%	-2.5%	+0.0%	+3.3%	+6.4%	+8.2%	+10.1%	+11.5%
MC-cBCI	5%	+1.4%	+6.1%	—	—	—	—	—
	10%	+2.7%	+5.5%	+9.8%	+12.8%	—	—	—
	15%	+2.7%	+7.3%	+10.7%	+13.3%	+17.1%	—	—
	20%	+3.0%	+7.6%	+10.4%	+12.4%	+13.7%	+15.2%	—
	25%	+3.5%	+8.0%	+11.4%	+14.7%	+17.4%	+20.5%	+22.2%
	100%	+3.9%	+8.9%	+12.3%	+15.6%	+17.7%	+20.1%	+21.5%
LDA-cBCI	5%	+0.2%	+4.1%	—	—	—	—	—
	10%	+3.3%	+5.8%	+9.2%	+12.5%	—	—	—
	15%	+4.0%	+7.6%	+10.2%	+12.4%	+18.3%	—	—
	20%	+4.0%	+8.1%	+10.5%	+12.3%	+13.9%	+16.1%	—
	25%	+4.4%	+9.9%	+13.8%	+16.6%	+19.5%	+21.9%	+23.7%
	100%	+4.7%	+11.2%	+14.8%	+17.1%	+19.1%	+20.1%	+21.5%

consulted in Section C.2.

Interestingly, the highest improvements for the MC-cBCI and LDA-cBCI approaches are not obtained for low values of δ . On the contrary, they occur when the threshold of acceptance to form groups is increased.

6.3.2.3 Comparison with $\max(AUC_1, AUC_2, \dots, AUC_r)$

We will now study the effects of placing ourselves in the second reference system, and comparing the performance of the cBCI with that of the best performer within the group. Table 6.8 reports, for each method, the median gains at difficulty level 2 for each group size r and different values of δ .

The results for the SC-cBCI method differ from those observed in the target

Table 6.8: Median improvements over the best participant in the group when using collaborative BCIs for target localisation at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size						
		2	3	4	5	6	7	8
SC-cBCI	5%	<i>-12.8%</i>	-11.4%	—	—	—	—	—
	10%	-11.0%	-10.0%	-9.0%	-4.1%	—	—	—
	15%	-12.7%	-10.9%	-10.5%	-10.0%	-13.2%	—	—
	20%	-12.6%	-10.5%	-10.1%	-7.7%	<i>-4.4%</i>	-4.0%	—
	25%	-12.5%	-11.3%	-10.5%	-9.1%	-8.0%	-6.8%	-7.1%
	100%	-12.6%	-11.5%	-10.2%	-8.2%	-7.0%	-5.9%	-5.4%
MC-cBCI	5%	<i>-1.3%</i>	+2.7%	—	—	—	—	—
	10%	-1.9%	+0.5%	+3.0%	+3.9%	—	—	—
	15%	-2.5%	-1.4%	+0.6%	+2.3%	+3.8%	—	—
	20%	-2.5%	-1.6%	-0.5%	+1.6%	+2.7%	+2.8%	—
	25%	-3.9%	-2.1%	-1.2%	+0.0%	+1.4%	+1.7%	+2.4%
	100%	-4.2%	-2.5%	-1.3%	+0.1%	+1.4%	+2.1%	+2.8%
LDA-cBCI	5%	-1.5%	+0.9%	—	—	—	—	—
	10%	-1.1%	+0.6%	+1.8%	+3.6%	—	—	—
	15%	-2.4%	-0.9%	+1.1%	+3.7%	+4.8%	—	—
	20%	-2.8%	-1.5%	+0.5%	+1.4%	+2.4%	+3.1%	—
	25%	-2.8%	-1.2%	+0.5%	+1.3%	+2.1%	+2.8%	+2.9%
	100%	-2.7%	-1.1%	+0.4%	+1.3%	+2.0%	+2.6%	+3.0%

detection system: Table 6.8 shows that this approach is, in almost every case, worse than the best performer of the group. However, the MC-cBCI and LDA-cBCI methods behave similarly to each other and to the results obtained for the target detection system. Moreover, these collaborative forms with participant selection outperform the cBCI without it. The highest gains for the MC-cBCI approach are obtained for intermediate values of the dissimilarity index (i.e., $\delta \in \{10, 15\}\%$), whereas in the case of LDA-cBCIs, they are obtained for the smallest value, $\delta = 5\%$ ¹.

Hence, in the case of multiple classifier systems (MC-cBCI and LDA-cBCI),

¹Although Table 6.8 only contains data for group sizes of 2 at the most restrictive value of δ , this comment is based on the behaviour of the system for other levels (see Appendix C).

cBCIs tend to produce better AUCs than the “intelligent” reference system too. Moreover, as opposed to the target detection system, the collaborative BCI still outperforms the best participant even when no participant selection is applied and all groups are allowed (i.e., $\delta=100\%$).

The results observed here are generalised across all difficulty levels for the target localisation cBCIs, as can be seen in Section C.2.3.

The improvements achieved using the MC-cBCI approach over the best performer of the group, across all levels and for different group sizes are summarised in Figure 6.7. We can see that the biggest improvements are concentrated at the bottom of each plot, indicating again a tendency of cBCIs to outperform the best participant when the members of the group have similar AUCs (i.e., low standard deviations of the AUC across the group).

The figure also highlights the fact mentioned above that bigger groups lead to higher improvements. As we saw from the results in the tables, even though for small groups it is possible to perform worse than the best participant of the group, this scenario is less common for bigger group sizes.

6.3.3 Theoretical Model of the Participant Selection Method

This section introduces a simple model to explain the possible reasons for the further improvements in performance that are obtained when groups are formed taking into account the similarity of the individuals.

The AUC for each participant can be interpreted as a measure of how spread and separated the distributions of scores for each class are. The bigger the overlap in these distributions, the lower the AUC value and *vice versa*.

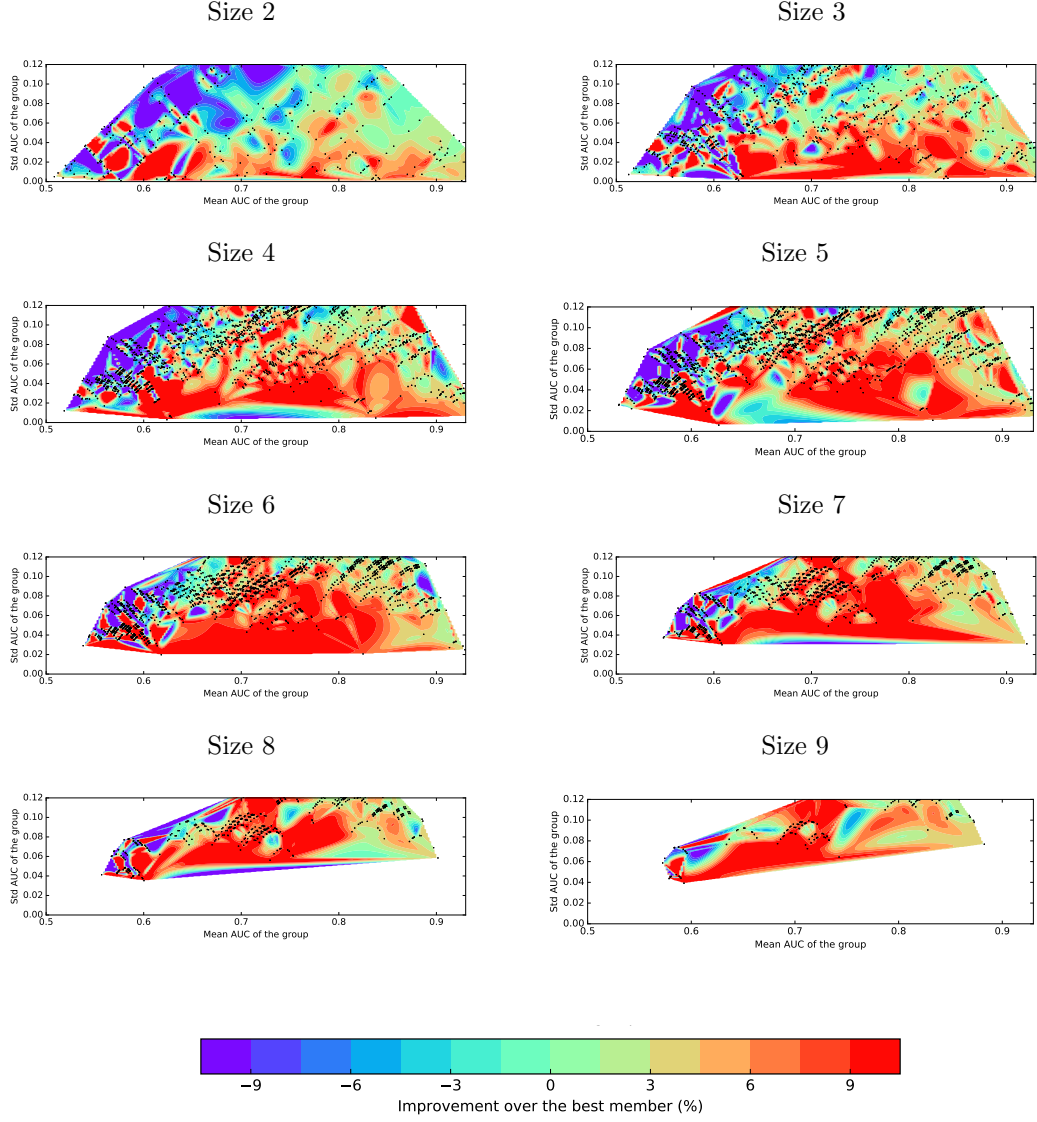


Figure 6.7: Surface interpolation of the percentage of AUC improvements over the AUC of the best member of the group using MC-cBCI for groups of size 2–9.

The MC-cBCI method consists of averaging classifiers' outputs to obtain the AUC of the cBCI for each group of participants. If we first focus on groups of size 2, the distribution of the average of two *uncorrelated* stochastic variables is the convolution of their pdfs (save for a scaling factor). Formally, let $S_{i,c}$ be

a stochastic variable representing the scores produced by a classifier for class $c \in \{C_1, C_2\}$ ¹ and participant $i = 1, \dots, 11$, and let $\text{pdf}_{i,c}(x)$ be its probability density function. Then, the pdf of the average of the scores for participants i and j when presented with a stimulus of class c , $S_{i,j,c} = (S_{i,c} + S_{j,c})/2$, is given by $\text{pdf}_{i,j,c}(x) = (\text{pdf}_{i,c} * \text{pdf}_{j,c})\left(\frac{x}{2}\right)$, where $*$ is the convolution operator.

For simplicity, let us assume that the variables $S_{i,c}$ are normally distributed, i.e., $S_{i,c} \sim \mathcal{N}(\mu_{i,c}, \sigma_{i,c}^2)$. Because the convolution of two Gaussians is a Gaussian, we have that also $S_{i,j,c} \sim \mathcal{N}(\mu_{i,j,c}, \sigma_{i,j,c}^2)$ with

$$\mu_{i,j,c} = \frac{\mu_{i,c} + \mu_{j,c}}{2} \quad \text{and} \quad \sigma_{i,j,c}^2 = \frac{\sigma_{i,c}^2 + \sigma_{j,c}^2}{4}.$$

Let us further assume that all participants have the same means for the two classes, i.e., $\mu_{i,C_1} = \mu_{C_1}$ and $\mu_{i,C_2} = \mu_{C_2}$, for $i = 1, \dots, 11$, and that the standard deviations for the classes are identical, i.e., $\sigma_{i,C_1} = \sigma_{i,C_2} = \sigma_i$ (but not the same for each participant). In this case, we have that

$$\mu_{i,j,c} = \mu_c \quad \text{and} \quad \sigma_{i,j,c}^2 = \sigma_{i,j}^2 = \frac{\sigma_i^2 + \sigma_j^2}{4}.$$

That is, the mean becomes independent from the pair (i, j) that forms the group, and the standard deviation is independent from the class, but depends on the (i, j) pair.

The separation between the distributions of scores jointly produced by a pair of participants can then be compared with the separation between the distributions of scores of the better performer from the pair. To do this, given the

¹Here, C_1 and C_2 could be classes L and R in LVF vs RVF classification, or T and NT in the case of target detection.

aforementioned assumptions, only the group's variance, $\sigma_{i,j}^2$, needs to be compared against the variance of the better participant of the group, which can be obtained as $\sigma_{min}^2 = \min(\sigma_i^2, \sigma_j^2)$. Since the means of the distributions for classes C_1 and C_2 remain constant, the AUC (calculated from the pdfs of the distributions) of the group will be better than that of the better participant when $\sigma_{i,j}^2 < \sigma_{min}^2$. If we estimate the parameters of the distributions (i.e., means and standard deviations) from real data, the theory presented here allows computing the AUCs for all cases, so it is possible to compute the expected gains/losses.

Figure 6.8 shows the expected gains of pairs over the better participant of the pair predicted by this model under the assumptions above. The parameters for the Gaussian variables used in the simulations (i.e., $|\mu_{C_1} - \mu_{C_2}| = 1$ and standard deviations $\sigma_i \in [0.3, 4]$) were estimated from the data collected from the RSVP experiment, for the left vs right classification task.

The general trend using the proposed model is that gains are higher (with respect to the AUC of the better participant) when the participants are very similar (i.e., similar AUCs, represented at the bottom of the figure). Even though one can find differences when comparing the theoretical improvements in Figure 6.8 with the actual results in Figures 6.2 and 6.7(a), the general similarity between the figures is striking, suggesting that a significant proportion of the effect is captured by the model.

Under the assumptions listed above, the model can easily be generalised to groups of size r . In this case, the distributions of scores for a group, for each class, i.e., $S_{R,c} \sim \mathcal{N}(\mu_{R,c}, \sigma_{R,c}^2)$, are determined by parameters

$$\mu_{R,c} = \mu_c \quad \text{and} \quad \sigma_{R,c}^2 = \sigma_R^2 = \frac{\sum_{i \in R} \sigma_i^2}{r^2},$$

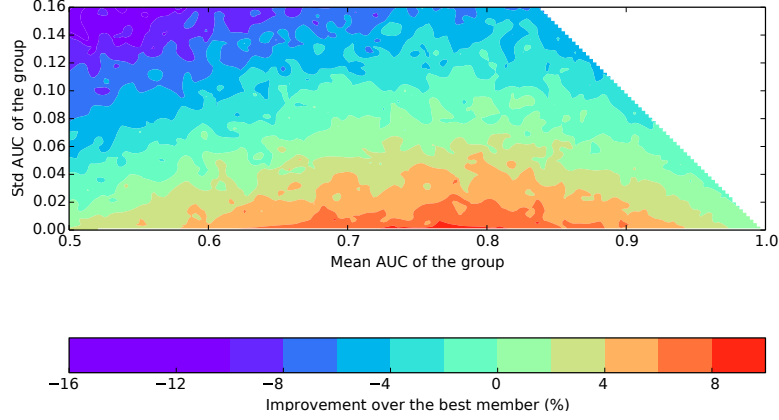


Figure 6.8: Expected gain of the joint AUC over the better participant of a pair when the distributions of scores for both classes are given by normally distributed random variables, $S_{i,c} \sim \mathcal{N}(\mu_c, \sigma_i^2)$, with $|\mu_{C_1=L} - \mu_{C_2=R}| = 1$ and standard deviations $\sigma_i \in [0.3, 4]$.

where R is the set of r participants included in the group. As before, in this case, the AUC resulting from the groups' scores for each class will be higher than that of the best participant if $\sigma_R^2 < \sigma_{min}^2$, with $\sigma_{min}^2 = \min_{i \in R}(\sigma_i^2)$.

Figure 6.9 shows the expected gains for groups of different sizes over the best participant of each group predicted by this model under the assumptions above. The figure illustrates the same trend as before, and also adheres to the experimental results from previous sections.

6.4 Discussion

This chapter introduced one of the core contributions of this thesis. In particular, a method for selecting members to form a group based on how similar the performance of the BCIs controlled by different individuals are was presented. The similarity measure was the absolute difference between the maximum and the minimum AUCs of the candidate group members. This method was applied

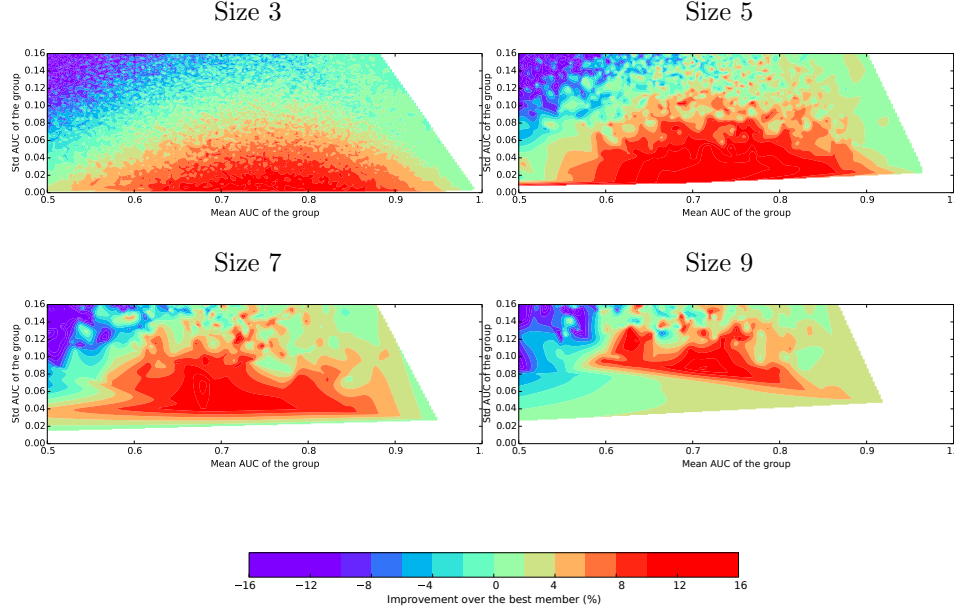


Figure 6.9: Surface interpolation of the expected improvements in the AUC (in percentage) over the AUC of the best member of the group, for different group sizes, according to the theoretical model, with the same parameters used in Figure 6.8.

to the target detection and target localisation systems, and compared the results obtained with two artificial systems: one that represents the average group performer and a second one that behaves like the best member of the team.

Even though performance of cBCIs is typically higher than corresponding single-user BCIs, performance increases dramatically when only participants with relatively close performance participate in a group. Under the assumption that the AUCs from single-user BCIs are correlated to the sensitivity of the individual's visual system, this result is in agreement with those of Bahrami *et al.* [2010] in their visual perception experiment: pairs of participants perform better than individual observers, provided that they had similar visual sensitivities.

Moreover, this result was consistent across the three approaches of cBCI used

in this thesis. When all participants are given equal weight in the cBCI (i.e., the SC- and MC-cBCI approaches), for high values of the dissimilarity-index δ , good performers are dragged down by those participants who did not perform so well. However, the LDA-cBCI approach is a more intelligent method of combining each participant's contributions, as it assigns them different weights for voting. For this reason, it is possible to still outperform the best member of the group even for high values of δ .

A theoretical model was developed that explains some of the reasons for the higher improvements in the performance of the cBCI systems when groups are formed taking into account the similarity of the individuals in the group. Even though the model used for the simulations did not reflect all the complexity of the system and made several assumptions about the underlying distributions of scores, the theoretical and empirical results are very similar.

Lastly, we looked at the distributions of individual AUCs from the N2pc and P300-based systems to understand the reasons behind the drop of performance in the N2pc system for pairs of participants observed in Section 4.3.4, and compared the results of the selection mechanism for these systems.

6.5 Conclusions

This chapter introduced the last aspect that we wanted to explore with the RSVP experiment. In it, we looked at the reasons behind the improvements of our collaborative BCI systems and developed a theoretical model that can explain part of those improvements.

Over the last few chapters, we have considered multiple aspects of the com-

bined RSVP–BCI paradigm, and explored some novel aspects of this technique.

In the next chapter we will consider a new scenario. Movies can be considered to be a special case of the RSVP paradigm, in which the images are very correlated with each other and presented at rates of above 25 Hz (and they usually include a sound track). However, in contrast to the bursts of images that are typically presented in the RSVP paradigm, movies tell a story, and in order to do this, editors and film makers influence the emotion of the viewer through features such as motion and brightness [Detenber *et al.*, 1998; Simons *et al.*, 1999]. Perhaps for all these reasons, the information presented through film is easier to remember than that of static images [Candan *et al.*, 2016]. Cuts have been related to increases of attention, which in turn are correlated with an increase of memory retention for the information provided after the occurrence of the cut [Lang *et al.*, 1993; Rothschild *et al.*, 1986].

Movies produce more robust neural responses than static images [Marathe *et al.*, 2016]. For this reason, we decided to use feature movies in the next step of this thesis, and to do a cut-by-cut analysis of the neural response that is elicited by cuts.

Chapter 7

A Preliminary Study of Cuts in Movies Using Collaborative BCI Techniques

Movies are sequences of related images that are displayed at a high frequency. They can, thus, be considered as a special case of the RSVP technique (plus sound). This chapter focuses on the changes that arise in the brain activity as a result of perceiving a cut in motion pictures (i.e., movies). This is done through an ERP analysis that uses techniques borrowed from the SC-cBCI method and a study of the event-related synchronisation/desynchronisation in different frequency bands.

7.1 Introduction

Movies consist of sequences of related images that are typically presented at frequencies between 25–30 Hz, which allow us to perceive them as continuous. In this sense, they can be considered as a very fast RSVP paradigm (plus the addition of sound).

In these sequences, editors introduce cuts (i.e., discontinuities) in order to create narrative effects. Since the main topic of this thesis until now has been a study of the brain’s response to the appearance of a target in RSVP streams of non-continuous images, we will now explore the effect of perceiving a discontinuity in a rapid stream of continuous images (that is, a cut).

As shown in Section 2.5, electroencephalography and other brain imaging techniques, such as fMRI, have been used to study the effects of movies on the viewer’s brain to some extent (e.g., [Hasson *et al.*, 2004, 2008]). In particular, some work using EEG has been done to study the brain’s response to the appearance of cuts.

Cuts induce alpha blocking on viewers, and the magnitude of this effect has been correlated with retention of content of television commercials [Appel *et al.*, 1979; Lang *et al.*, 1993; Reeves & Thorson, 1986; Reeves *et al.*, 1985; Rothschild *et al.*, 1986; Smith & Gevins, 2004]. The time to recovery to normal values of alpha may be taken as a measure of how well attention is held by the content. Rothschild *et al.* [1986] showed that not all scene changes are strong enough to elicit alpha blocking, and, if they do, the rate of recovery to previous levels of power in the alpha band may depend on how well the scene holds attention of the viewer. This may be connected with the concept of *edit blindness*, a phenomenon

by which some cuts go completely unnoticed by viewers [Smith & Henderson, 2008b]. The probability of detecting a cut seems to vary depending on factors such as the type of cut and the amount and direction of motion and audio before and after it [Smith, 2012; Smith & Yvonne Martin-Portugues Santacreu, 2016].

Anderson *et al.* [2006] performed, perhaps with the exception of Zacks *et al.* [2010] (see below), the only comparison between related and unrelated cuts using fMRI. They found that most of the brain activations that occurred in response to a cut were common despite of the type of cut, with the exception of an area which has been associated with the processing of information to create meaning.

In addition to fMRI and studies of the EEG in the frequency domain, since cuts are very sharp in the time domain, they have also studied, although to a lesser degree, through ERPs which are time-locked to the occurrence of the cut [Francuz & Zabielska-Mendyk, 2013]. In this area of research, cuts have been associated with slow cortical potentials, which are also used as an index of the orienting response, and other ERPs [Francuz & Zabielska-Mendyk, 2013], such as the (frontal) SNW1, (parietal) SPW and (parietal) SNW1. When comparing the ERPs that are produced in response to related vs unrelated cuts, Francuz & Zabielska-Mendyk [2013] found that unrelated cuts produce more negative SNW1 and more positive SPW responses than those produced by related cuts. The lack of elicitation of a P300 ERP was also noted by the authors [Francuz & Zabielska-Mendyk, 2013].

Finally, in movie stimuli in which congruent or incongruent information appeared following a cut, an N400 was detected over frontal, central and parietal regions [Reid & Striano, 2008]. This component was followed by a late large positive potential in posterior regions of the brain in the sequences with an in-

congruity [Sitnikova *et al.*, 2003, 2008], perhaps an indication that this unexpected turn of events triggers other cognitive processes in addition to the ones that mediate pure semantic integration.

While these results indicate that cuts produce characteristic ERPs, these have always been studied in either averages or grand averages (i.e., after averaging hundreds of individual responses across several participants). Moreover, to the best of our knowledge, no ERP analysis of the neural response has been performed using feature movies: the work surveyed until now involved either commercial messages, audiovisual content created specifically for the experiment, or short fragments from television shows.

We wondered to what degree it would be possible to detect the ERPs induced by cuts in movies on a cut-by-cut basis through some form of BCI. We also wanted to perform an analysis of the ERPs induced by cuts in feature movies, something that had not been done before, to the best of our knowledge. Of course, the real-world nature of the stimuli (as well as the fact that people watching movies move, laugh, blink, etc.) makes this a very hard task.

It was discussed previously in this thesis that, due to the noise affecting EEG recordings, in BCIs it is common to perform ERP averages across a small set of repetitions of the same stimuli (typically 3–6 repetitions) in order to increase the signal-to-noise ratio, and correspondingly achieve better classification. However, there are situations where this technique cannot be used. This is certainly the case when studying the brain’s response to specific events in movies, because the reaction to the first presentation of an event is often different from those produced in further repetitions of the event [Dorr *et al.*, 2010; Nittono, 2008]. For this reason, in this study we borrowed the *technique* used in the SC-cBCI approach of

the collaborative BCI of *averaging EEG signals from multiple participants on a trial-by-trial basis* in order to increase the SNR. The reason behind choosing the averaging technique as used in SC-cBCI, as opposed to alternatives such as the LDA-cBCI, which was shown to give a superior performance in previous chapters, is that the study presented this chapter there is not a BCI in the strict sense. We aim at presenting a study of cut-by-cut ERPs, rather than trying to classify or distinguish among them.

Moreover, previous EEG and fMRI studies have shown that there is a high inter-subject correlation in participants' signals while they watch movies [Bridwell *et al.*, 2015; Hasson *et al.*, 2008], so averaging across multiple participants should indeed help to increase the SNR.

We hoped that, thanks to the analyses performed and presented in this chapter, it would be possible to make inferences about how and what type of processing happens across a cut. This can be studied by measuring the coherence before, after and across cuts between different brain regions [Tucker *et al.*, 1986], or through other measure of brain functional connectivity [Daly *et al.*, 2012].

Regardless of the fact that no actual BCI is directly derived in this work, the results presented in this chapter may be used to inform and build a cBCI in the future. Such a cBCI could be used to study the effectiveness of a cut (or some other aspect of the movie) in the viewer. Practical applications could include selecting which movie trailer is expected to cause an intended effect and, therefore, produce the highest revenue, or deciding where and/or how to place a cut while editing the film.

This chapter begins by explaining the methods used for conducting this investigation (Section 7.2), including a description of the protocol for data collection

and the signal processing techniques applied to the EEG recordings. Section 7.3 presents the results of the analysis, which are then discussed on Section 7.4, where also some suggestions for future work are provided.

7.2 Methods

7.2.1 Data Acquisition

Data were collected from 10 volunteers (aged 21–40, mean age = 28.7 years old, 2 females, 1 left handed) with normal or corrected-to-normal vision. One participant was excluded from the study due to technical difficulties during data recording. The remaining participants (N=9) had a mean age of 29.6 years (range: 25–40, 2 females, 1 left-handed). They all signed the consent form approved by the Ethics Committee of the University of Essex.

Participants were seated at approximately 80 cm from the 20-inch LCD screen where the stimuli were presented. EEG data were acquired by using a BioSemi ActiveTwo system with 64 electrodes mounted in a standard electrode cap following the international 10-20 system. The EEG was referenced to the mean of the electrodes placed on the earlobes. The initial sampling rate was 2048 Hz. Data were band-pass filtered between 0.15 and 40 Hz before downsampling to 128 Hz.

Correction for eye blinks and other ocular movements was performed by applying the same subtraction algorithm that was used in previous chapters, based on correlations with the average of the differences between channels Fp1 and F1 and channels Fp2 and F2 [Quilter *et al.*, 1977] (see Section 3.2.1.1).

Table 7.1: Details of the movie clips used in the experiment.

<i>Movie</i>	<i>ID</i>	<i>Beginning</i>	<i>Duration (sec)</i>	<i># cuts used</i>
Iron Man 1 (IM1)	IM1.1	0:02:16	83	34
	IM1.2	0:38:34	137	49
	IM1.3	1:17:16	124	43
Iron Man 3 (IM3)	IM3.1	0:22:45	61	13
	IM3.2	0:25:29	51	13
	IM3.3	0:34:09	330	133
	IM3.4	1:41:59	846	330
Sherlock Holmes (SH)	SH	1:23:30	152	17
The Expendables (Exp)	Exp1	0:43:08	89	42
	Exp2	1:22:15	429	179
V for Vendetta (Ven)	Ven1	0:10:18	46	20
	Ven2	2:01:55	155	57

7.2.2 Stimuli

Participants watched 12 movie clips extracted from 5 feature movies. Details for each clip are given in Table 7.1.

The clips were displayed in full-screen mode at a 1680×1050 resolution, subtending 30.29×19.23 degrees of visual angle. Audio was played by means of a pair of high quality desktop speakers (Edirol Roland MA-15D). The average sound level measured from the participant’s position (approximately 2 m from the sound sources) with an Arco CR:262 device was 55.1 dBA, with a maximum level of 64.6 dBA (the reference sound level in the quiet room was 30.7 dBA).

All participants watched the clips in the same order. At the end of a clip, participants could press a mouse button to start the next video.

Cuts were detected manually and the programme kept a list that contained the last frame of each scene. In order to guarantee synchronisation between the video and the EEG, the clips were played using the API for the VideoLAN player, which can be used to print out the current frame. When the frame being played

by the VideoLAN player was identified as a cut (by comparing the frame number to that in the list), the script sent a trigger to the EEG system.

Since the aim of this work is to study ERPs that are spontaneously elicited in response to cuts, participants were asked to simply watch the movies, without performing any specific task.

At the end of the experiment participants were asked to rank the clips in order of preference.

7.2.3 Signal Processing

7.2.3.1 Epoch Extraction

For ERP analysis, 1000 ms-long epochs were extracted from the EEG signals, time-locked to the occurrence of a cut. All epochs were referenced to the mean of the 200 ms interval preceding them.

In order to avoid contamination, for the “cut” condition, only epochs belonging to a cut that was not followed by another cut within 1000 ms were considered throughout the whole analysis. The first scene of each clip was also discarded. This resulted in 930 “cut” epochs (from the original 1429 cuts) per participant.

Moreover, in order to establish a baseline for determining the brain’s response to cuts, epochs for a “non-cut” condition were also considered. “Non-cut” epochs consisted of the 1000 ms immediately preceding a cut. Again, in the interest of minimising the effect of ERPs from previous cuts, only epochs where there had been no other cuts within 1500 ms were considered. This resulted in 566 “non-cut” epochs.

Epochs for which the amplitude was greater (in absolute value) than $200\mu V$

at any electrode were removed from the analysis (an average of 11 cuts per participant). In total, 8271 cut epochs were analysed.

Given previous results found in the literature (see Section 2.5), we were not expecting P300s to appear in response to cuts. However, we investigated whether the N2pc appears in response to one, signalling a shift of attention towards either side of the frame. To the best of our knowledge, this had not been attempted before. For this, epochs were extracted from electrodes PO7 and PO8 in the time interval comprising the 1000 ms immediately following cuts that were expected to induce a shift of attention to the left or right side (as manually labelled by a pool of volunteers — more on this below) were included.

In addition to the ERP analysis, we decided to complement the study with a frequency analysis, as is common in this type of studies. For this, two different methods were used: first, we looked at standard event-related synchronisation and desynchronisation (ERS/ERD) [Kalcher & Pfurtscheller, 1995; Klimesch *et al.*, 1998; Pfurtscheller & Lopes da Silva, 1999] plots to validate and compare our results to those from the literature (e.g., the study of alpha blocking in response to a cut [Appel *et al.*, 1979; Lang *et al.*, 1993; Reeves & Thorson, 1986; Reeves *et al.*, 1985; Rothschild *et al.*, 1986; Smith & Gevins, 2004]). Secondly, we went one step further and studied the connectivity networks that are formed in the brain in response to a cut, in an attempt to connect the changes observed in the frequency domain with theories of context-updating and working memory. This was done through the measure of the coherence between different brain regions.

The frequency analysis considered the following frequency bands: theta (4–8 Hz), lower alpha (8–10 Hz), upper alpha (10–14 Hz), low beta (14–18 Hz) and upper beta (18–26 Hz).

7.2.3.2 Event-Related Synchronisation and Desynchronisation

ERP analyses are useful to detect and study time-locked changes in EEG activity, such as those who occur in response to sensory stimuli (e.g., the P300 or the N2pc). However, these and other stimuli can also cause alterations in the EEG which are not phase-locked, and, thus, cannot be extracted through averaging [Pfurtscheller & Lopes da Silva, 1999]. Such frequency-specific changes typically consist of power increases or decreases in the ongoing EEG, which in turn are considered to be due to an increase or decrease in the synchrony of the underlying neuronal populations.

In order to study the ERD/ERS, signals are first band-passed filtered in the frequency band of interest. This was done in this study through a 4th order Butterworth filter. All samples are then squared to obtain power samples, and the average is calculated across all trials and a time window (in this study, the window was of 50 ms), so a final value is obtained for each electrode site considered.

Furthermore, to discriminate between phase-locked and non-phase locked power changes (i.e., to eliminate the possible influence of ERP components), the point-to-point intertrial variance can be calculated instead of squaring the time samples [Kalcher & Pfurtscheller, 1995; Klimesch *et al.*, 1998]:

$$IV_t = \frac{1}{N-1} \sum_{i=1}^N (x_{i,t} - \bar{x}_t)^2,$$

where N is the number of trials, t represents the time sample, $x_{i,t}$ represents the amplitude of the EEG signal for trial i at time t and \bar{x}_t is the mean of the data (across all filtered trials) at time t .

Finally, the ERD is calculated as the percentage change of the power or in-

tertrial variance (A_t) at each time point (or an average of time samples) with respect to the power in a reference period (R , calculated over the time interval $[n_0, n_0 + k]$). By using the formula

$$ERD(\%) = \frac{A_t - R}{R} \cdot 100,$$

with

$$R = \frac{1}{k} \sum_{t=n_0}^{n_0+k} A_t,$$

positive (resp. negative) values represent synchronisation (resp. desynchronisation) and an increase (resp. decrease) of power in the band considered.

In this study, we used the non-phase locked calculation for the ERD/ERS (i.e., the intertrial variance, rather than the instantaneous power). The reference period was the interval $[-300, 0]$ ms referred to the onset of a cut.

7.2.3.3 Spectral Coherence

Spectral coherence (or just coherence, from here onwards) is defined as the cross-correlation of the power spectrum of two signals, normalised by their power spectra (i.e., the normalised cross-spectral density):

$$C_{xy} = \frac{|P_{xy}|^2}{P_{xx} \cdot P_{yy}}$$

Coherence provides evidence for the functional connectivity or coupling between different brain areas during cognitive processing [Sauseng *et al.*, 2005; Tucker *et al.*, 1986; Weiss & Rappelsberger, 1996]. It can therefore be interpreted

as a measure of information transfer between the areas under consideration [Weiss & Rappelsberger, 1996].

Coherence values range between 0–1, with 0 meaning (in our context) that there is no functional connectivity between the areas/electrodes considered.

Again, the frequency bands considered in the coherence analysis were theta (4–8 Hz), lower alpha (8–10 Hz), upper alpha (10–14 Hz), low beta (14–18 Hz) and upper beta (18–26 Hz).

As in the ERP analysis, in an attempt to avoid contamination from previous cuts, only those cut/non-cut epochs that were far from other cuts were used for the analysis.

7.2.4 Left/Right Attention Shifts

The last frame before each cut and the first one after the cut were extracted and sequentially shown (in order) to 7 participants that matched the mean age of those who participated in the original EEG experiment.

For each image, volunteers were asked to manually select (by pressing a mouse button) the location of the image where they felt their eyes were first directed to, and to do so as quickly as they could.

The x- and y-coordinates selected by each participant (starting at the top left of the image and with increasing x and y values towards the bottom and right side of the screen, respectively) for each image were stored and analysed as follows: first, we used two methods to assess the degree of concordance between different participants. These were the Overall Concordance Correlation Coefficient (OCCC) [Barnhart *et al.*, 2002; Lin *et al.*, 2002; Lin, 1989] and the Normalised

Scanpath Saliency (NSS) [Dorr *et al.*, 2010], which are explained below.

Secondly, after seeing the large variance of the NSS results, only those cuts for which there was a high synchrony both in the last frame *before* and the first frame *after* the cut were considered. Since we wanted to keep the dataset as clean as possible, we decided to set a high threshold for deciding whether there was high attentional synchrony or not. In particular, we chose the third quartile across all the NSS values obtained for all the images.

Then, the attentional spots selected by the participants were averaged for each image to obtain the x-coordinate of the attentional spot.

Finally, since we were only interested in studying those cuts for which a shift of attention was expected, only those cases where the difference between the x-coordinates of the attentional spots before and after the cut were ≥ 111.11 pixels (i.e. 2° of visual angle, which is approximately the area covered by the fovea [Luck, 2012]) were considered. This was decided based on the assumption that the eyes of the viewers would be fixated on the attentional spot of the frame before the cut and thus no attention shift would happen if the attentional spots before and after the cut were too close to each other on the horizontal axis.

7.2.4.1 Overall Concordance Correlation Coefficient

The concordance correlation coefficient was first proposed by Lin [1989] as a measure of agreement between two different methods or judges. For example, it could be used to validate a new method by comparing it to an established golden rule.

The general case used when several methods are compared to each other is called the Overall Concordance Correlation Coefficient (OCCC) and expressed in

terms of the means, variances and covariances of the different judges across all measurements, is given by the following equation [Barnhart *et al.*, 2002]:

$$\rho_c = \frac{2 \sum_{j=1}^{J-1} \sum_{k=j+1}^J \sigma_{jk}}{(J-1) \sum_{j=1}^J \sigma_j^2 + \sum_{j=1}^{J-1} \sum_{k=j+1}^J (\mu_j - \mu_k)^2},$$

where J is the number of judges, μ_j represents the mean for judge j , σ_j^2 is the variance of judge j , and σ_{jk}^2 is the covariance between the pair of judges j and k .

The OCCC ranges between -1 and 1, where the minimum value represents perfect discordance between the judges and 1 is complete agreement between them.

7.2.4.2 Normalized Scanpath Salience

The NSS was first developed to compare how saliency models match human gaze data. However, it can also be used to assess inter-subject variability [Dorr *et al.*, 2010].

In order to compute the NSS, a fixation map needs to be constructed for each image. Instead of using features extracted from the images themselves to create such maps, we used the coordinates chosen by the other participants to construct a fixation map in a leave-one-subject-out loop. This allowed us to evaluate the similarity of the left-out participant to all the other observers.

The method to construct the fixation maps is as follows [Dorr *et al.*, 2010]. For each image i and observer j from the training set S , the x coordinate of the attentional spot x_j^i was obtained. For each of them, a spatial Gaussian G_j^i centered around this coordinate was placed in a spatial fixation map F . The Gaussians

were added across the training set participants for each frame. Formally,

$$F(x) = \sum_{j \in S} \sum_{i=1}^M G_j^i(x)$$

where M is the total number of images and

$$G_j^i(x) = e^{-\frac{(x - x_j^i)^2}{2\sigma_x^2}}.$$

F was normalised to have zero mean and a standard deviation of 1. This is what is called an NSS map.

$$N(x) = \frac{F(x) - \bar{F}(x)}{SD(F(x))}$$

Finally, the NSS for the left-out participant is computed as the average of the NSS maps evaluated at the coordinate chosen by the test participant x_k^i .

$$NSS = \frac{1}{M} \sum_{i=1}^M N(x_k^i)$$

The interpretation of this measure is as follows: positive values of the NSS indicate that the test participant is choosing attentional spots in agreement with those chosen by the participants used to generate the maps. If, however, the chosen spots are uncorrelated with those from the rest of the group, the NSS will be 0. Lastly, negative values indicate that the attentional regions chosen by the test participant are very different from those of the training group.

After assessing the similarity between the participants, we used the data from all the participants to generate a new NSS for each frame, and the maximum

Table 7.2: Range of ratings and median ranking given by participants to each movie clip. Low numbers indicate higher preference. Rows have been sorted by the median of the rankings (last column).

<i>Movie ID</i>	<i>Range</i>	<i>Median</i>
Ven2	[1, 5]	2.0
Ven1	[1, 4]	3.0
IM1.1	[1, 6]	4.0
IM3.3	[1, 8]	4.5
IM3.4	[3, 7]	5.0
IM1.2	[3, 9]	6.5
IM3.1	[6, 11]	7.5
IM3.2	[7, 10]	8.0
IM1.3	[7, 12]	9.0
SH	[2, 10]	10.0
Exp1	[4, 12]	11.0
Exp2	[10, 12]	11.5

value was used to determine whether there was high attentional synchrony or not. In particular, we used the third quartile as a threshold to decide whether the frame would be considered for further analysis or not.

7.3 Results

7.3.1 Behavioural Results

Table 7.2 summarises the rankings for each movie clip taking into account the responses from all participants.

In general, the participants agreed on their assessments of the different clips. For example, the least preferred clips belonged to the movie “The Expendables” for all except one participant. Similarly, 9 out of 10 participants ranked Ven2 with a 1 or a 2, and the clip from Sherlock Holmes with a 10. Opinions were

divided across all clips from Iron Man 1 and Iron Man 3, which could add noise to the results for these two movies when using this information in the ERP analysis below if clip preference is related in any way to the ERPs that are evoked in response to cuts.

7.3.2 ERP Analysis

Figure 7.1 shows the time course of the grand average of the ERPs at several central and centro-posterior channels, for the cut and non-cut conditions. A large negative ERP starting 100 ms after the cut and extending until after 700 ms is clearly present in the cut condition, but not in the non-cut condition. The peak amplitude of this component decreases (in absolute value) as we move toward the posterior electrodes.

This ERP has a similar time course to the frontal SCPs shown by Francuz & Zabieliska-Mendyk [2013]. However, the waveform that they reported was positive toward the back of the head (e.g., Pz), whereas the one we found remains negative throughout.

The other main ERP study using cuts that we are aware of is that of Sitnikova *et al.* [2008]. The time course and scalp topography of our ERP is quite similar to theirs, specially for frontal electrodes.

The scalp maps representing the scalp distribution of voltages at 200 ms and 400 ms after a cut shown in Figure 7.2 offer evidence that this component is present in frontal and central electrodes, extending also to centro-posterior sites¹

¹This ERP is unlikely to be associated with eye blinks because these would manifest as large positive deflections at frontal electrodes with much reduced effects at central and posterior sites (neither of which is the case in Figure 7.1), and we have corrected for ocular movements and eye blinks.

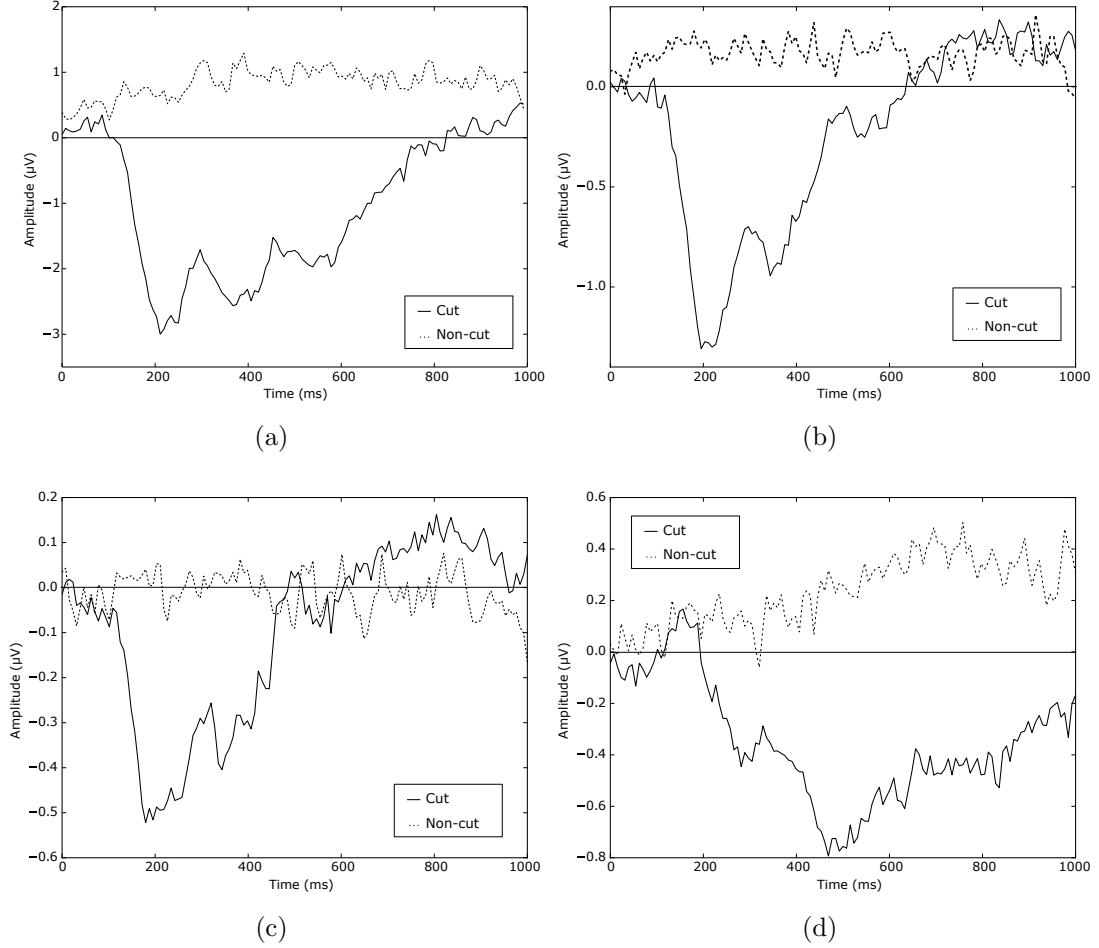


Figure 7.1: Grand average plot for cuts and non-cuts at electrodes (a) FCz, (b) CPz, (c) Pz, and (d) Oz.

In order to derive a more detailed, cut-by-cut, analysis, we then used the cBCI technique of averaging signals across the 9 participants. For each epoch in the averaged signals, the median amplitude was calculated across the time interval 380–420 ms, where the second peak of this ERP appears, across all electrodes. We term this value *Post-Cut Negativity* (PCN) hereafter. Median amplitudes of the PCN for cuts were larger (in absolute value) than median amplitudes for non-cuts (-1.77 and $0.09 \mu V$, respectively). The distributions in the two groups

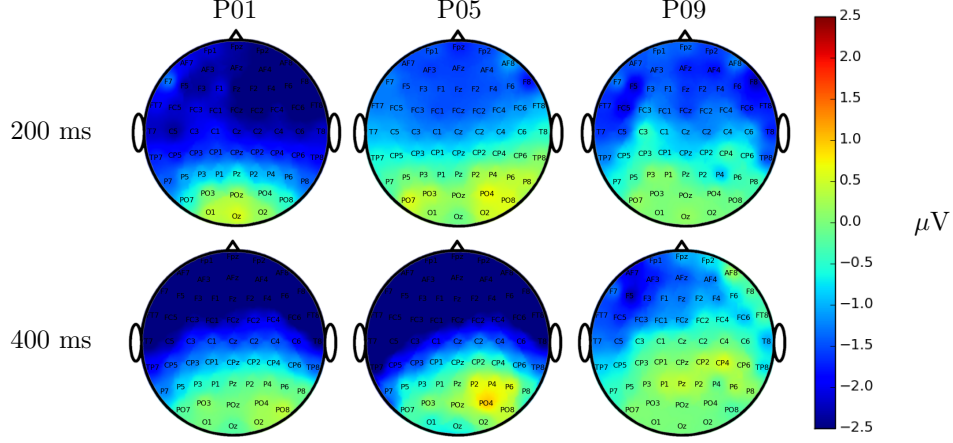


Figure 7.2: Scalp distributions of voltage amplitudes 200 and 400 ms after a cut for three representative participants.

Table 7.3: P values from a two-sided Mann-Whitney test with Bonferroni correction for PCNs across different movies.

	<i>IM3</i>	<i>SH</i>	<i>Exp</i>	<i>Ven</i>
<i>IM1</i>	0.39	4.9×10^{-2}	5.2×10^{-3}	1.9×10^{-2}
<i>IM3</i>	—	1.7×10^{-2}	1.1×10^{-2}	5.7×10^{-2}
<i>SH</i>	—	—	1.5×10^{-3}	3.1×10^{-3}
<i>Exp</i>	—	—	—	0.74

differed significantly ($W = 369,415$, $p < 2.2 \times 10^{-16}$) according to a two-sample Mann-Whitney test comparing cut and non-cut conditions.

7.3.3 Influence of Movie and Cut Characteristics on PCN

As illustrated in Table 7.3, which reports the Bonferroni-corrected p values obtained from a two-sided Mann-Whitney U test comparing PCNs across different movies for the cut condition, PCNs vary across movies.

We studied the effects of clip preference (i.e., median subjective ranking of a clip) and median PCN amplitude in that clip. However, no correlation between these factors was found (Spearman's $\rho = -0.25$, $S = 358.4$, $p = 0.21$).

We then explored whether shot length, cut luminance and optical flow could be factors in determining such differences in PCNs. Results of these analyses are reported in the following subsections.

7.3.3.1 Shot Length

We studied the effects of shot length (defined as the period between two cuts) on PCN amplitude. Figure 7.3 shows that the median shot length and the median PCN for different clips and movies are correlated (Spearman's $\rho = -0.7$, $S = 486.3$, $p = 5.6 \times 10^{-3}$). Due to noise, the PCN for *individual cuts* shows a smaller correlation with the length of the shot, but this is highly statistically significant (Spearman's $\rho = -0.23$, $S = 1.7 \times 10^8$, $p = 2.4 \times 10^{-13}$). These results indicate that the PCN is related to scene length, with long scenes resulting in larger (in absolute value) PCNs. Since only those cuts that were not preceded by another one within the previous second were not included in the analysis, this effect is unlikely to be caused by shorter cuts being baselined at a lower activity level as not having time to settle back to 0.

7.3.3.2 Luminance

Luminance was calculated as in Cutting *et al.* [2012]: the last frame preceding each cut and the first after it were first converted to grayscale, which results in values between 0–255. The luminance for each image was then calculated as the median for that frame.

No correlation was found between the average luminance of the first frame following a cut and PCN (Spearman's $\rho = 0.02$; $S = 1.3 \times 10^8$, $p = 0.57$), nor between the average luminance of the last frame before the cut and PCN

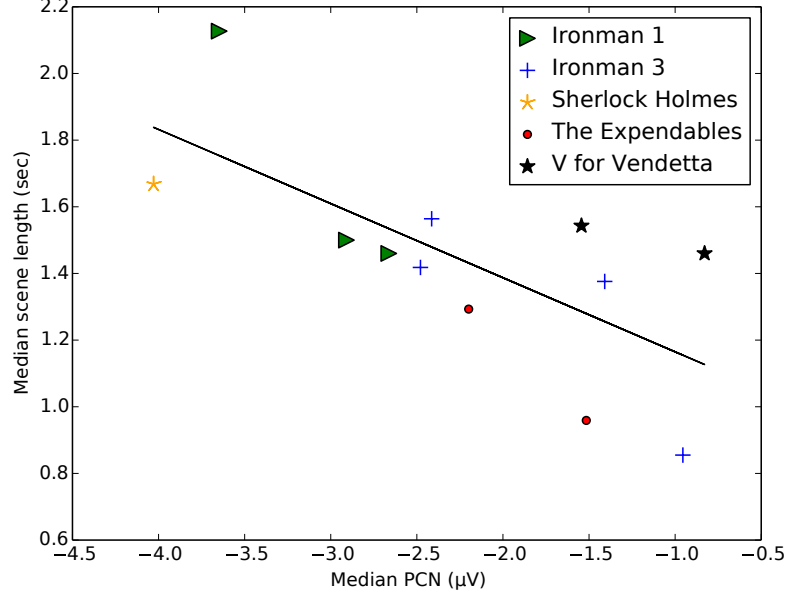


Figure 7.3: Median shot length as a function of median PCN.

(Spearman's $\rho = 0.07$; $S = 1.2 \times 10^8$, $p = 0.03$).

Finally, the difference in luminance across the cut was calculated as the median luminance after subtracting each value from the frames following and preceding each cut. Again, no correlation was found between this factor and PCN amplitude (Spearman's $\rho = 0.13$; $S = 1.3 \times 10^8$, $p = 0.13$). Similarly, These results indicate that variations of PCN amplitude are not related to the low-level features of a scene.

7.3.3.3 Optical Flow

The optical flow (or image velocity) is a measure of the distribution of apparent velocities of movement of brightness patterns in a sequence of images. It is an estimate of the shift that is needed in an image in order to obtain the next one. It can, thus, be used to determine whether there is motion before, after or on the

two sides of a cut.

In order to check the impact of motion on the PCN, the average optical flow was calculated across the ten frames preceding a cut, the average optical flow across the ten frames after the cut and the difference between them through Farnebäck [2003]’s algorithm as implemented in the OpenCV library Bradski [2000].

There was no correlation between any of these and the amplitude of the PCN associated to that cut (Spearman’s $\rho = -0.05$, $\rho = -0.02$ and $\rho = 0.02$, respectively). Hence, there is no evidence that motion before, after or across a cut has an impact on the amplitude of the PCN evoked by the cut.

7.3.4 Shifts of Attention in Movies

As pointed out previously, researchers that looked into the brain’s response to cuts found differences between related and unrelated cuts, but no P300-like ERPs [Francuz & Zabielska-Mendyk, 2013], so we did not expect to find this component in our data. However, to the best of our knowledge, no attempt at finding the N2pc to signal a shift of attention after a scene change has occurred.

We will first look at the behavioural response of the participants that annotated the focus of attention for each frame surrounding a cut, and will then show the ERP (calculated as in previous chapters, by means of the contralateral and ipsilateral waveforms at channels PO7 — located on the left hemisphere — and PO8 — on the right on selected cuts).

Table 7.4: Average normalised scanpath salience (NSS) and overall concordance correlation (ρ_c) coefficient for each video clip.

<i>Movie ID</i>	<i>NSS</i>	<i>ρ_c</i>
IM1.1	1.67	0.33
IM1.2	1.62	0.16
IM1.3	1.53	0.26
IM3.1	1.67	0.24
IM3.2	1.53	0.26
IM3.3	1.67	0.22
IM3.4	2.01	0.33
SH	1.96	0.40
Exp1	1.78	0.21
Exp2	1.36	0.23
Ven1	1.86	0.56
Ven2	1.80	0.47
Average	1.72	0.30

7.3.4.1 Inter-Subject Agreement in Attentional Spots

In order to evaluate the quality of the ground truth provided by the volunteers when selecting the focus of attention, we took into account the attentional synchrony that is expected from edited films [Smith & Henderson, 2008a; Smith & Mital, 2013].

Let us start by looking at the degree of synchrony between the participants. Table 7.4 shows the average NSS and OCCC from the leave-one-participant-out loop for each of the video clips. Across all participants and movies, the average NSS was 1.72, and the OCCC was 0.3.

One of the advantages of the OCCC with respect to the NSS is the fact that its value is bound between ± 1 , so there exist guidelines for its interpretation. In general, values of the OCCC between 0.4–0.59 are typically considered to signal fair agreement between the judges [Cicchetti, 1994]. In comparison, the values

that we obtained are quite low, and only three of the clips belong to In contrast, our average is below this number, and only three of the clips shown to participants have a ρ_c in this range. Hence, we decided not to use all the data available, but only those for which there was a very high NSS value.

If f denotes the first frame after the cut, we asked participants to annotate the first part of the scene to which they were drawn to, for frames f and $f - 1$, for all cuts from all the video clips that the first group of participants had watched. The attentional spot for each image was calculated as the mean of all the x-coordinates determined by the participants.

The attention shift was then calculated as the difference in the x-coordinate of the attentional spot at frame f minus that of frame $f - 1$.

Cuts were divided into those that should have evoked a shift of attention to the left and those that were expected to shift the viewers' attention to the right.

7.3.4.2 ERPs in Attention Shifts

Due to the very high restrictions that we imposed on our dataset, the final number of cuts for which we expected an attentional shift (signalled by the N2pc) was much lower than the original number of cuts. In total, there were 48 cuts belonging to the “left” and 21 cuts in the “right” categories. Hence, a total of 621 cuts are included across all participants.

Figure 7.4 shows the grand average of the contralateral minus the ipsilateral waveforms recorded at electrode sites PO7 and PO8 for cuts that were expected to produce a shift of attention towards the left or right side of the screen.

Although a lot of noise is present in this grand average, an N2pc-like ERP is found starting around 250 ms after the cut. We believe the reason for this noise

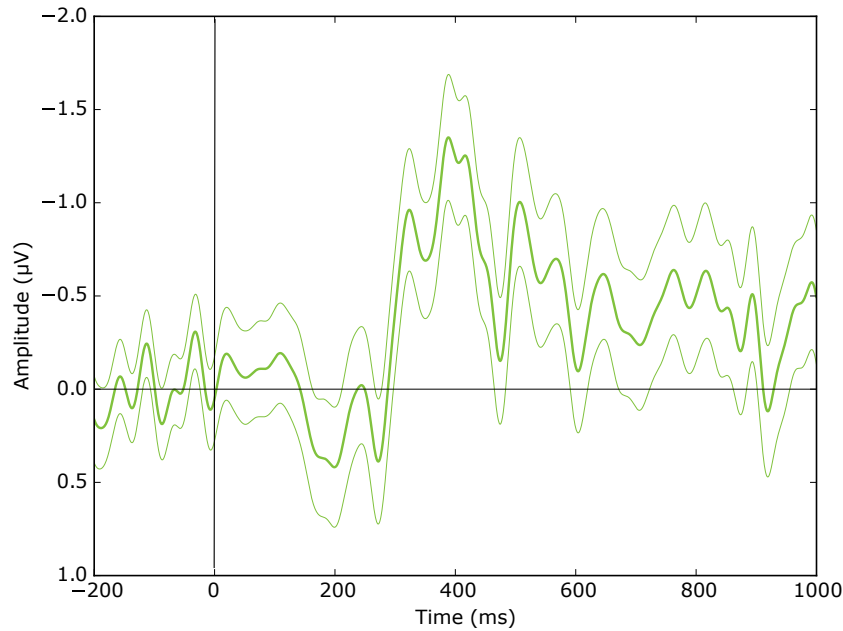


Figure 7.4: Contralateral minus ipsilateral grand averages for left and right cuts across all participants.

is that the shift of covert attention is not produced when the cut occurs, but rather, as suggested by Smith [2012], viewers may have been cued before the cut is introduced in order to make the cut invisible, so the shift of attention occurs before. If this was the case, given the fact that we are locking the epochs to the occurrence of the cut and the actual shift of attention occurs at a variable time before it, we would not be able to obtain an ideal-like N2pc.

In any case, at this point, we cannot be certain that the ERP component from the figure is indeed an N2pc.

7.3.5 Frequency Analysis

Previous analyses of variations in the EEG while watching video segments found in the literature focused mainly on the frequency domain. Such previous work reported variations of power in the alpha band (e.g., [Reeves *et al.*, 1985]), which were then related to changes in mental effort and attention. For example, it has been shown that cuts evoke a decrease of power in the alpha band (also known as *alpha blocking*), which is known to be representative of a higher mental load or to a focusing of attention [Lang *et al.*, 1993; Reeves & Thorson, 1986; Reeves *et al.*, 1985; Rothschild *et al.*, 1986; Smith & Gevins, 2004].

In this chapter we looked at variations of EEG activity in the frequency domain, through the traditional study of event-related synchronisation and adding an analysis of changes in coherence between pairs of electrodes, which, to the best of our knowledge, had not been attempted before.

7.3.5.1 ERD/ERS Analysis

ERD/ERS maps allow us to look at the distribution of the evoked response to the occurrence of a cut in different frequency bands. The average ERD/ERS calculated across all participants and movie clips are shown for different time intervals and frequency bands (as described on page 183) in Figure 7.5.

These plots show the alpha blocking effect that had been previously reported. Indeed, the only frequency band for which an ERD effect is clear is the upper alpha band, whereas the rest of the bands show ERS patterns across the whole scalp, specially in the low beta frequency range.

Despite the fact that these grand average plots seem to be an all-or-nothing

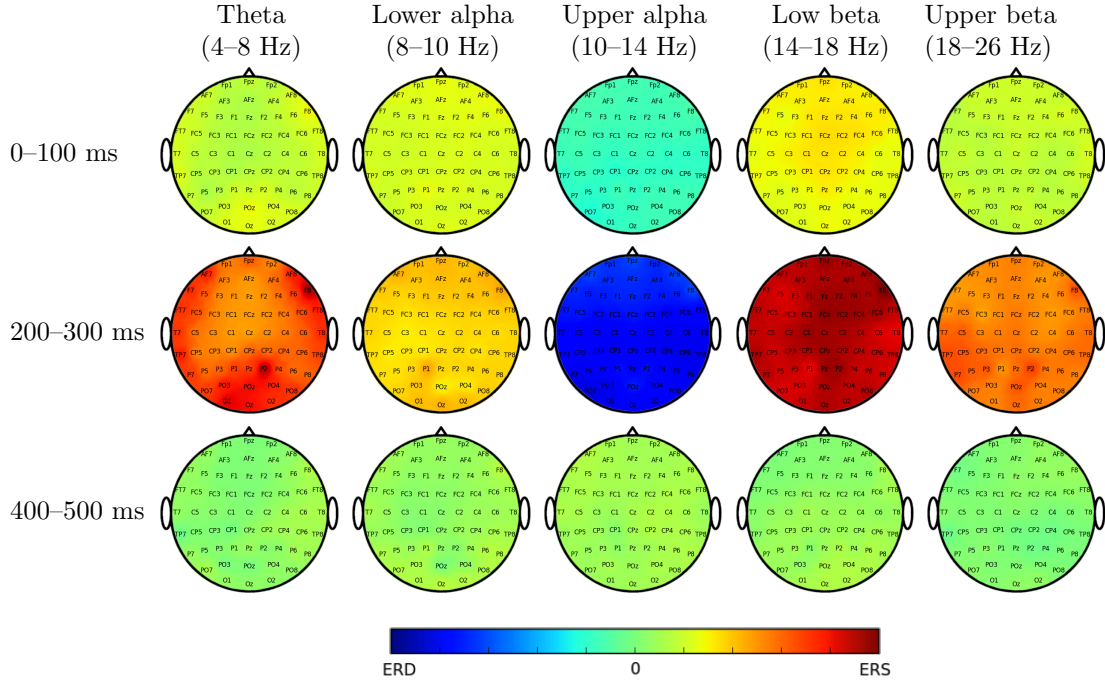


Figure 7.5: Grand averaged spatial distribution of ERD/ERS after the occurrence of a cut at different time intervals (from top to bottom: [0,100], [200,300] and [400,500] ms, referred to cut onset), for theta, lower alpha, upper alpha, low beta and upper beta bands.

response, individual ERD/ERS plots show more variation on the scalp, as shown in Figure 7.6, which represents the ERD/ERS response to cuts for participant 003. Even though the color scale is the same as for Figure 7.5, much more detail can be appreciated from the individual ERD/ERS scalp plots when compared to the grand averages. As we noted in Section 2.2.1.3, grand averages show the earliest onset and latest offset of a component [Luck, 2014; Rothschild *et al.*, 1986]. These may vary noticeably across participants, as seems to be the case for the ERD/ERS plots. However, the grand average only shows the most persistent features, and not those that are not common to all participants.

The individual ERD/ERS plots for the remaining participants can be seen in

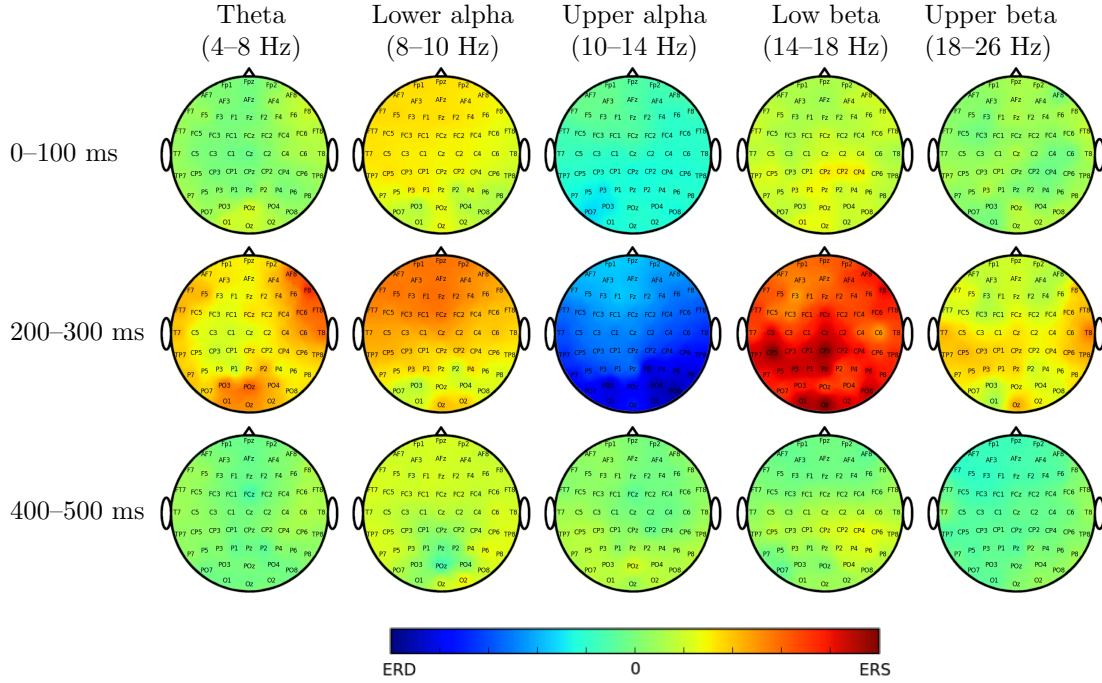


Figure 7.6: Spatial distribution of ERD/ERS after the occurrence of a cut for participant 003 at different time intervals (from top to bottom: [0,100], [200,300] and [400,500] ms, referred to cut onset), for theta, lower alpha, upper alpha, low beta and upper beta bands.

Appendix D. Even though they, in general, display a higher variance in colours with respect to the grand average, some patterns are common to all participants: the very high level of desynchronisation in the upper alpha band that is spread along the scalp, and similarly high level of synchronisation in the beta band (both low and upper beta ranges). These are consistent with the literature and, in particular, with the phenomenon known as alpha blocking [Appel *et al.*, 1979; Lang *et al.*, 1993; Reeves & Thorson, 1986; Reeves *et al.*, 1985].

So far we have looked at ERD/ERS plots that show the behaviour in the frequency domain of each electrode site in response to cuts. However, this type of analysis does not take into account the way in which different regions interact

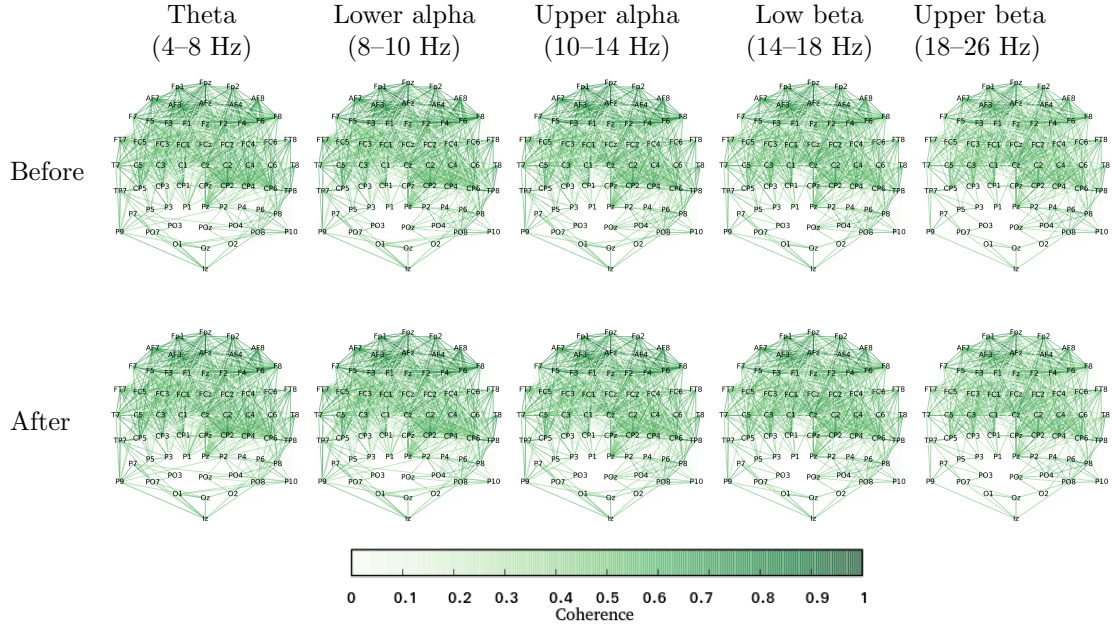


Figure 7.7: Average coherence between pairs of electrodes regions before (top row) and after (bottom row) the occurrence of a cut in different frequency bands.

or influence each other. For this, we looked at the coherence between all possible pairs of electrodes, before, after and across cuts.

7.3.5.2 Variations in Spectral Coherence Evoked by Cuts

Figure 7.7 shows the average coherence (across all participants and movies) between all possible pairs of electrodes before and after cuts for different frequency bands. For clarity purposes, it should be noted that only those links between electrodes for which the coherence was greater than 0.45 were plotted in this figure. For the interested reader, the original coherence plots are shown in Appendix D.

There are no major changes between the “before” and “after” conditions for a given frequency range, a fact that indicates that the functional coupling (or connectivity patterns) between brain areas is not affected by the cut when the

same time intervals are considered (i.e., when only epochs from before or only epochs from after the cut are used). In fact, coherence either stays the same or slightly decreases after the cut (with respect to before the cut). This drop is most evident in the upper beta band. This seems to indicate that the occurrence of a cut does not significantly alter the information transfer between brain areas.

Moreover, the highest coherence is consistently found in the anterior regions of the brain across all frequency bands, both before and after the cut, and decreases as we move towards the back of the scalp, with the lowest values being located in the posterior areas (especially in the left hemisphere). These findings are consistent with previous reports of coherence during different cognitive tasks (e.g., [Rappelsberger & Petsche, 1988; Sauseng *et al.*, 2005]).

However, when coherence is calculated across the cut, the pattern changes, as shown in Figure 7.8. Coherence for a given region calculated between pairs of regions using epochs before and after the cut decreases to values between 0.29 and 0.32 in all electrode pairs and frequency bands considered. Given the small range of values for each topoplot, no threshold for the minimum value of coherence was applied to these plots.

The results shown in this figure, and in Figure 7.9, which illustrates the coherence for different frequency bands between the pre- and post-cut conditions at each particular electrode, show that the dominant frequency within each of the bands considered changes after the occurrence of the cut. Moreover, they also highlight that there are changes in information transfer across the brain regions that occur after the occurrence of a cut.

The fact that the functional connectivity between pairs of electrodes before and after the cut is not affected (as shown in Figure 7.7 means that the cut does

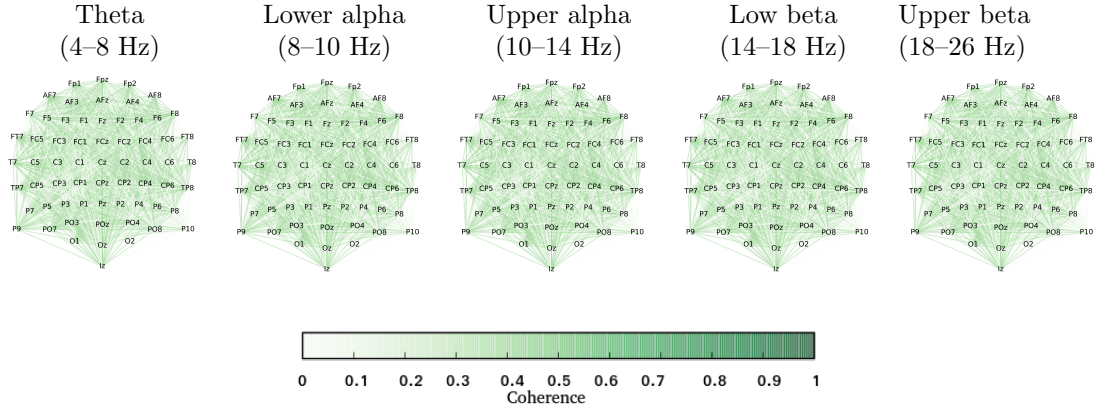


Figure 7.8: Average coherence between pre- and post-cut conditions for pairs of electrodes.

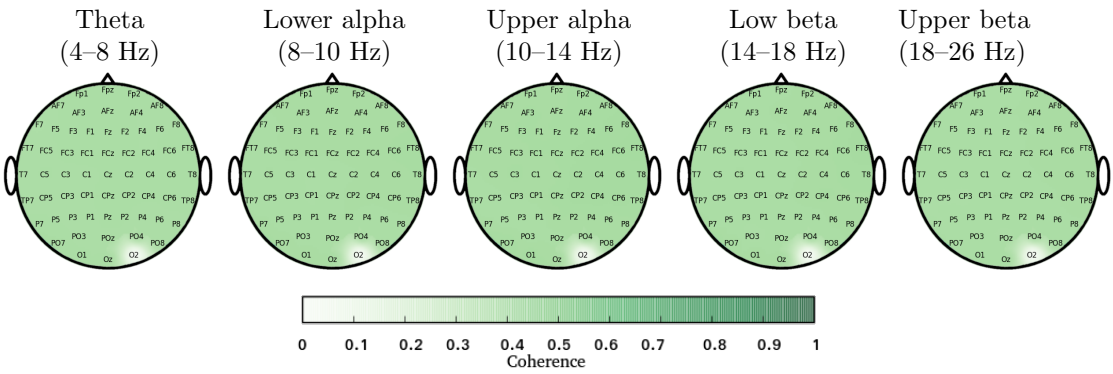


Figure 7.9: Average coherence between pre- and post- epochs for each electrode.

not alter the information transfer between different parts of the brain. However, the drop in coherence seen in Figures 7.8–7.9 indicates that the dominant frequency within each band changes. As we discuss below, this is consistent with the ERD/ERS plots from the previous section (Figure 7.5), which showed power changes in all frequency bands after the occurrence of a cut.

7.4 Discussion

This chapter explored what happens in the brain as a result of cuts in feature movies. In particular, both the time (through an ERP analysis) and the frequency domains (including ERD/ERS and coherence across different regions in the scalp) were considered.

A large and distinctive negative ERP was found time-locked to the occurrence of scene changes. This ERP is distributed on most frontal and central electrode sites and on several parietal sites in most participants. This ERP could be either an N400 or the SCPs identified in previous research using more constrained types of videos Francuz & Zabielska-Mendyk [2013]; Reid & Striano [2008]; Sitnikova *et al.* [2003, 2008]. In particular, our ERP seems to fit with the time course and scalp topography of the N400 component reported by Sitnikova *et al.* [2008]. These components are linked to processes taking place during the integration of new semantic information acquired after the cut into a context built before the cut.

Because this ERP has such a wide scalp distribution and an approximate latency of 400 ms, we chose to characterise it by using the median voltage across all channels in a 380–420 ms time window. Also, to further increase the SNR of the ERP, thus allowing a cut-by-cut analysis in the time domain, we borrowed the technique of averaging epochs (on a trial-by-trial basis) across participants from the area of cBCIs. The resulting value was called the post-cut negativity.

While the PCN is clearly a very coarse-grained quantity along the spatial and participant axes, it is still maximally sharp in the time domain. That is, each cut has an associated PCN value. This makes it possible to attempt, for the first

time, to characterise cuts at a finer resolution than that of the standard taxonomy including only related and unrelated cuts.

In particular, we were able to highlight a previously unreported relationship between shot length and the corresponding PCNs in related cuts. This may be due to movies with longer scenes having richer plots, resulting in more complex context updates being required following rarer cuts. In contrast, fast-paced movie clips showing rapid sequences of cuts (e.g., due a number of related explosions) would only require smaller context updates (and associated PCNs) after the first cut (perhaps an unexpected explosion). It should be noted that, even in the case of movies with short scenes, only ERPs that were free of baseline contamination from previous cuts were included. Thus, it is unlikely that these results are due to issues with the baseline process.

This is somehow consistent with the behavior of other ERPs, such as the P300, the amplitude of which is modulated by the rarity (and, correspondingly, the waiting time) of target stimuli [Polich, 2004a]. However, unlike P300 amplitudes, PCNs do not depend on low-level features such as the changes in luminance necessarily associated with cuts, further indicating the high-level nature of the cognitive processes associated with cuts and represented by the PCN.

In the analysis, it was also found that PCNs varied across movies and, to a lesser extent, clips extracted from the same movie. This makes sense, as different techniques for joining shots may be used by professional movie editors, based on their own style, experience and the intended effect of the cut.

Particularly appealing from this perspective is the possibility of building collaborative BCIs, perhaps based on finer features of cut-related ERPs than the simple PCN used in this chapter. These could *objectively* evaluate whether spe-

cific cuts in a movie actually achieve the effect intended by the movie editor or director, during the test screening phase of the movie. For example, as shown in Section 2.5, one of the fields that the attentional theory of cinematic continuity [Smith, 2012] the reasons behind some cuts being perceived by viewers whilst other are invisible (also termed edit blindness) [Smith & Henderson, 2008b]. In this work we did not ask participants to detect cuts, so it is not possible to correlate the amplitude of the PCN with the probability of a cut being missed. However, recent research shows that motion before and/or after the cut (e.g., in match-action types of edits, where a cut is introduced after the onset of motion and shows the continuation of that motion or action in the second shot) affects this probability [Smith & Yvonne Martin-Portugues Santacreu, 2016]. Unfortunately, the basic experiment included in this chapter that considered the optical flow as a measure of movement in a scene did not show relationships between PCN amplitude and movement before and after the cut. However, the analysis did not distinguish amongst the many different types of cuts and did not consider which proportion of these were of the match-action type, so it is not possible to directly compare these results with the aforementioned literature.

It is a well-known fact that power in the upper alpha (especially in the left hemisphere) and theta bands is linked to the processing of semantic information [Klimesch, 1996, 1999]. Given that theta synchronisation signals memory retrieval and encoding of information, the ERS found after the cut in this frequency band might indicate the activation of context-updating processes. In this light, the synchronisation observed in theta (as well as the alpha blocking in the upper alpha region — associated to long-term memory processes) will also support the hypothesis that the PCN is signalling the integration of new seman-

tic information [Klimesch, 1996, 1999; Sauseng *et al.*, 2005; Schack *et al.*, 2005] mentioned above.

Lastly, we looked at the spectral coherence between pairs of electrodes before, after and across the occurrence of a cut. The results show that coherence is stronger between anterior brain areas than between posterior regions, both before and after the cut. Taken together with the results above, and, in particular, with the hypothesis that the PCN and the changes in ERD/ERS that we found signal a context updating process in the brain, the fact that coherence is higher in frontal regions of the brain is consistent with the hypothesis that the prefrontal cortex is highly involved in working memory processes [Sauseng *et al.*, 2005], e.g., for updating the previously-built context with the new information presented after the cut.

Moreover, coherence decreases greatly across all pairs of regions when epochs before and after the cut are considered. Since coherence is the cross-correlation of two spectra, the observed drops in coherence values indicate changes in the dominant frequencies within each band. This is related to the synchronisation and desynchronisation of the frequency bands considered: synchronisation in a given band implies that more groups of neurons start oscillating at frequencies that are very close to each other, thus limiting the bandwidth of the oscillation frequencies, increasing the chance of interference being constructive (i.e., the sum of the oscillations is larger) and, thus, the power in the band. In contrast, when desynchronisation occurs, oscillations at different frequencies might result in destructive interference, which in turn decreases bandpower.

7.5 Conclusions

In this chapter we studied the brain's response to cuts in action movies from Hollywood. We found a large ERP that is elicited by cuts and whose amplitude is not correlated with low-level features of the movie, which suggests that it may be related to processing of information and/or context updating processes. In addition to this ERP, we showed that an N2pc-like ERP is elicited in response to lateral shifts of attention in the cut. We cannot be sure at the moment that this ERP is indeed an N2pc component, but believe that it is quite likely to be one. Thus, we plan to further study this aspect in the future.

In addition to the ERP analyses, we performed an ERD/ERS analysis that showed that cuts in our experiment have the same alpha blocking effect that had been reported previously in the literature. We have not studied this effect on a movie-by-movie or cut-by-cut basis, but this is another aspect that we would like to pursue in the future.

Even though this chapter has barely scratched the surface of studying the effects of cuts on memory and perception, we feel that this research has opened up a number of promising avenues both for the psychophysiology of motion pictures and for applications of cBCIs in cinematography, television, advertising, etc. Particularly promising results may be obtained in the area of edit blindness, which might, in turn, help understand different mechanisms of the brain associated with memory and the processing of information.

An important fact that should be born in mind is that in this work we did not use EOG electrodes to monitor eye movements from the viewers in the free watching task. It is important in this type of work to ensure that the ERPs that

we have found are not the product of horizontal saccades, which are not accounted for in the artefact correction method that we used. However, considering the lack of activation in the frontal electrodes near the eyes (e.g., AF7, AF8), we do not believe that this is the case. Moreover, the similarity of our ERP with those previously reported elsewhere [Francuz & Zabielska-Mendyk, 2013; Sitnikova *et al.*, 2008] also seems to suggest that this is a real effect, and not the product of horizontal eye movements. In any case, a further data collection with eye tracking is recommended and remains as a task for the future.

Chapter 8

Conclusions and Future Work

This chapter ends the thesis with a summary of the main contributions of the work and a description of different avenues for future exploration.

8.1 Summary of Contributions

Brain-computer interfaces have been developed for a few decades. At first, they were created with the intention of restoring communication capabilities for people with severe disabilities. However, advances in different aspects related to BCIs, such as signal processing, processor capabilities and different technological aspects have made it possible to conceive such systems for augmenting human capabilities, e.g., through collaborative BCI systems in which multiple users control a single device.

This new architecture is very useful at fighting one of the main problems that traditional single-user BCI systems face: the high amount of noise that is recorded by the electrodes, an issue that is especially severe when moving out of

the ideal laboratory conditions. Electromagnetic interference (e.g., in the form of a 50 or 60 Hz component) and muscular artefacts are some of the causes of contamination in the EEG signals, which difficult the discrimination task for a classifier trained to distinguish between a number of classes or states.

Another major source of noise is the background activity of the brain, which is also affected by distraction or tiredness of the BCI user. External stimuli can trigger neural responses (e.g., ERPs) that can be picked up by EEG equipment and used to control a BCI. However, the background activity that is always present in combination with any other source of contamination in the EEG can mask these responses in single-trial systems. In systems where speed is a key factor, or where presenting the same stimulus to a user multiple times to gather more information to aid classification is not feasible, performance of the BCI system can be increased by merging evidence from multiple users.

This idea is the basis of this thesis, which explored different alternatives for blending information from multiple users through ERP-based collaborative brain-computer interfaces.

We first studied three different types of cBCIs in a visual search task in which participants had to find targets within rapid streams of images. One of the main findings of this work was the **presence of the N2pc ERP in such a complex scenario with real-world imagery** and its **modulation in amplitude** due to the distance from the target to the centre of the screen. This allowed our systems to not only detect the presence of a target, but also to **identify the area of the image where it was located** using the amplitude of this evoked potential in response to lateralised targets. Moreover, we found that the **N2pc varied significantly depending on the handedness of the user**.

As expected, collaborative BCIs were able to outperform single-user BCI systems. Moreover, we presented a **user selection method for grouping users** which increased the performance of the cBCI even further, and managed to achieve median improvements above 20% with respect to the best user of a group.

In the second experiment, participants were asked to watch a series of clips extracted from Hollywood feature movies. We used the approach of averaging EEG signals from single trials across multiple users. This technique allowed us to look in detail at a **previously unreported ERP, the post-cut negativity (PCN) that occurs in response to a discontinuity** (or cut) in the movie. We were able to perform a **cut-by-cut analysis** and found that the amplitude of this ERP was not affected by the low-level features of the scene, and that it **correlates with the length of the scene in a movie**, possibly pointing at a **context-updating mechanism that is signalled by the PCN**.

This **context-updating theory was further supported through an analysis in the frequency domain** that showed changes in ERD/ERS in the upper alpha and theta bands in the same way in which they are known to be related to long-term and working memory, respectively.

To summarise, the main contributions of this work are the following:

- Studying three ways of combining evidence from multiple users for a target detection task.
- Showing that the N2pc is found in the very hard conditions of our experiments, and that it varies significantly according to the user's handedness.
- Attempting a target localisation task, both in single-user and collaborative BCIs, using only EEG signals (the N2pc).

-
- Developing and validating a theoretical model for explaining the improvements that are obtained in the collaborative BCI systems.
 - Performing a cut-by-cut analysis of ERPs elicited in response to movies, and showing that this ERP (possibly an N400) is not affected by low-level features of movies.

8.2 Future Work

The work presented in this thesis has opened up a number of additional research pathways that we believe are worthy of exploration in future research.

One of the limitations of my RSVP experiment was the use of a limited amount of different possible target templates (i.e., between one and three different planes, although randomly rotated), and the repetition of one of the types of airplanes in most of the levels. Indeed, we saw a drop of performance in difficulty levels where pictures were presented at the same speed but contained more types of targets, with respect to levels where only one type of target was used. There are multiple possible causes for this, but we believe the most likely one is the effect of intertrial priming (see Section 2.4.3.3) to the most frequent template.

The intertrial priming effect has been studied mainly with pop-out types of targets (e.g., red targets in a display of black and white distractors), and with relatively basic shapes (e.g., vertical and horizontal lines). However, to the best of our knowledge, it has not been studied with complex stimuli such as the ones used in this PhD thesis. We believe that further research on this effect would be beneficial for triage systems.

As discussed in Section 4.3.1, another limitation of this work is that, for very

practical reasons associated with how participants react to high-speed protocols, the order of the 5, 6, 10, 12 and 15 Hz conditions was not randomised across subjects. Obviously, this would be required for a more precise comparison across presentation rates. In future research we will address this issue by extending the pre-experiment practice sessions, e.g., by inviting participants twice: once for practice, and a second time, after they are rested again, for the real experiment. This will make it possible for participants to adapt to the speed of the protocol before the real experiment starts, thereby allowing a fully counterbalanced experimental design.

Although most of the difficulty levels used in this work include only one specific template of an airplane, we saw that, for the relatively high presentation rate of 10 Hz, the changes in peak amplitude of the N2pc are not significant when moving from this paradigm to one which contains multiple different airplanes. In future work, we will need to verify if these results still hold for different shapes of targets and types of images, to see to what extent it is possible to build BCIs that can be used for target detection and localisation across a range of target types.

Moreover, the discovery of the differences in the N2pc related to handedness should be taken into account in the future, both in the literature of the N2pc related to attentional processes, and for creating cBCIs (especially in the single classifier approach — SC-cBCI — i.e., when averaging EEG epochs across multiple users).

Another limitation of this work is the amount of data that was analysed in the first scenario. Multiple methods have been tested on the same dataset, which may have resulted in our methods overfitting to these data. Thus, as we have indicated throughout this PhD thesis, applying our methods to a completely unseen test

set is needed in order to fully validate the results presented here.

With respect to the second scenario, as stated before, we believe it should be possible to link the presence and amplitude of the PCN with the theory of cinematic continuity. Determining what are the differences in the brain's response between cuts that are and those that are not perceived by a viewer might give us an additional insight into the psychophysiology of movie-watching, with possible applications in the cinematic community. The correlations used in this thesis (see Chapter 7) considered either the whole clip or individual cuts to study the PCN. Instead, smaller fragments of the clip (e.g., of 5–10 seconds of duration) might increase the resolution of this measure. For example, this would allow researchers to study whether the PCN correlates with the emotional state of viewers.

Furthermore, the movies selected for this experiment were explicitly selected for the high number of cuts presented in them, due to the fast pace that is typical of action movies (further enhanced by the occurrence of explosions). An extension of this analysis to other types of movies might provide extra information on the behaviour of the PCN.

Finally, further work is needed in the field in terms of artefact correction mechanisms and a study of the ways in which movement artefacts really affect BCI performance. Considering that the cBCIs presented in this thesis are designed mainly with the able-bodied population in mind — although they can, of course, be used by people with disabilities—, positive results in this area would definitely open the door to the use of the EEG to study neural responses and automatic reactions without requiring participants to stay still for the duration of the experiment. “Normal” behaviour is affected by the fact that a person is being observed. This is also the case when the person under observation is wearing a

device to monitor different types of physiological data, as is the case of an EEG cap, and more so when he/she is asked to remain as motionless as possible. Thus, if the requirements that we impose on participants in experiments can be relaxed, they will allow a more natural experience for the user (which might be a desired effect, e.g., in the movie-watching scenario), and might provide researchers with additional information, e.g., from a psychophysiological point of view.

Moreover, research in this area could also be appended to the results presented from the cBCI systems in the early chapters of this thesis, making also EEG-based triage systems more attractive for intelligence analysts. As it is unlikely that all group members will perform movements simultaneously, this would be the case if the robustness to noise achieved by our systems is able to cope better with the motion artefacts.

Collaborative BCIs are a promising solution to many of the issues currently inherent in single-user BCI systems. The work presented in this thesis has explored some of the open problems in this blooming field of research, but we believe that it has also opened some promising avenues for future work, that we will hopefully aid to explore with contributions from other enthusiasts from the scientific community, that we hope to keep being a part of.

References

- ACQUALAGNA, L. & BLANKERTZ, B. (2011). A gaze independent spelling based on rapid serial visual presentation. In *Engineering in Medicine and Biology Society (EMBS), 33rd Annual International Conference of the IEEE*, 4560–4563, IEEE.
- ACQUALAGNA, L. & BLANKERTZ, B. (2013). Gaze-independent bci-spelling using rapid serial visual presentation (rsvp). *Clinical Neurophysiology*, **124**, 901–908.
- ACQUALAGNA, L., TREDER, M.S., SCHREUDER, M. & BLANKERTZ, B. (2010). A novel brain-computer interface based on the rapid serial visual presentation paradigm. In *Engineering in Medicine and Biology Society (EMBS), 32nd Annual International Conference of the IEEE*, 2686–2689, IEEE.
- AKBANI, R., KWEK, S. & JAPKOWICZ, N. (2004). Applying support vector machines to imbalanced datasets. In *European Conference on Machine Learning*, 39–50, Springer.
- ALLISON, B.Z., MCFARLAND, D.J., SCHALK, G., ZHENG, S.D., JACKSON, M.M. & WOLPAW, J.R. (2008). Towards an independent brain-computer interface using steady state visual evoked potentials. *Clinical neurophysiology*, **119**, 399–408.
- ANDERSON, D.R., FITE, K.V., PETROVICH, N. & HIRSCH, J. (2006). Cortical activation while watching video montage: An fMRI study. *Media Psychology*, **8**, 7–24.

REFERENCES

- ANDO, T., TANAKA, A., FUKASAKU, S., TAKADA, R., OKADA, M., UKAI, K., SHIZUKA, K., OYAMADA, H., TODA, H., TANIYAMA, T. *et al.* (2002). Pupillary and cardiovascular responses to a video movie in senior human subjects. *Autonomic Neuroscience*, **97**, 129–135.
- APPEL, V., WEINSTEIN, S. & WEINSTEIN, C. (1979). Brain activity and recall of TV advertising. *Journal of Advertising Research*, **19**, 7–15.
- ARIGA, A. & YOKOSAWA, K. (2008). Attentional awakening: Gradual modulation of temporal attention in rapid serial visual presentation. *Psychological Research*, **72**, 192–202.
- ASTOLFI, L., TOPPI, J., FALLANI, F.D.V., VECCHIATO, G., SALINARI, S., MATTIA, D., CINCOTTI, F. & BABILONI, F. (2010). Neuroelectrical hyperscanning measures simultaneous brain activity in humans. *Brain topography*, **23**, 243–256.
- AWNI, H., NORTON, J.J., UMUNNA, S., FEDERMEIER, K.D. & BRETL, T. (2013). Towards a Brain Computer Interface Based on the N2pc Event-Related Potential. In *6th Annual International IEEE EMBS Conference on Neural Engineering*, IEEE, San Diego (CA).
- BABILONI, F. & ASTOLFI, L. (2014). Social neuroscience and hyperscanning techniques: past, present and future. *Neuroscience & Biobehavioral Reviews*, **44**, 76–93.
- BABILONI, F., CINCOTTI, F., LAZZARINI, L., MILLAN, J., MOURINO, J., VARSTA, M., HEIKKONEN, J., BIANCHI, L. & MARCIANI, M. (2000). Linear classification of low-resolution EEG patterns produced by imagined hand movements. *IEEE Transactions on Rehabilitation engineering*, **8**, 186–188.
- BAHRAMI, B., OLSEN, K., LATHAM, P.E., ROEPSTORFF, A., REES, G. & FRITH, C.D. (2010). Optimally interacting minds. *Science*, **329**, 1081–1085.
- BARNHART, H.X., HABER, M. & SONG, J. (2002). Overall concordance correlation coefficient for evaluating agreement among multiple observers. *Biometrics*, **58**, 1020–1027.

REFERENCES

- BARTELS, A. & ZEKI, S. (2004). Functional brain mapping during free viewing of natural scenes. *Human brain mapping*, **21**, 75–85.
- BERGER, H. (1929). Über das elektrenkephalogramm des menschen. *European Archives of Psychiatry and Clinical Neuroscience*, **87**, 527–570.
- BIGDELY-SHAMLO, N., VANKOV, A., RAMIREZ, R.R. & MAKEIG, S. (2008). Brain activity-based image classification from rapid serial visual presentation. *IEEE transactions on neural systems and rehabilitation engineering*, **16**, 432–441.
- BIRISAN, M. & BELING, P.A. (2014). A multi-instance learning approach to filtering images for presentation to analysts. *Environment Systems and Decisions*, **34**, 406–416.
- BLASCO, J.S., IÁÑEZ, E., ÚBEDA, A. & AZORÍN, J. (2012). Visual evoked potential-based brain-machine interface applications to assist disabled people. *Expert Systems with Applications*, **39**, 7908 – 7918.
- BRADLEY, A.P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, **30**, 1145–1159.
- BRADSKI, G. (2000). The OpenCV Library. *Dr. Dobb’s Journal of Software Tools*.
- BRIDWELL, D.A., ROTH, C., GUPTA, C.N. & CALHOUN, V.D. (2015). Cortical response similarities predict which audiovisual clips individuals viewed, but are unrelated to clip preference. *PLoS ONE*, **10**, 1–20.
- BRUNNER, P., JOSHI, S., BRISKIN, S., WOLPAW, J.R., BISCHOF, H. & SCHALK, G. (2010). Does the “P300” speller depend on eye gaze? *Journal of Neural Engineering*, **7**, 056013.
- CANDAN, A., CUTTING, J.E. & DELONG, J.E. (2016). RSVP at the movies: dynamic images are remembered better than static images when resources are limited. *Visual Cognition*, 1–12.

REFERENCES

- CECOTTI, H. (2011). Spelling with non-invasive Brain-Computer Interfaces - Current and future trends. *Journal of Physiology - Paris*, **105**, 106–114.
- CECOTTI, H. & RIES, A. (2015). Implication of non-stationarity in single-trial detection performance of event-related potentials. In *15th UK Workshop on Computational Intelligence (UKCI)*, Exeter (UK).
- CECOTTI, H. & RIVET, B. (2014). Subject combination and electrode selection in cooperative brain-computer interface based on event related potentials. *Brain Sciences*, **4**, 335–355.
- CECOTTI, H., KASPER, R.W., ELLIOTT, J.C., ECKSTEIN, M.P. & GIESBRECHT, B. (2011a). Multimodal target detection using single trial evoked EEG responses in single and dual-tasks. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 6311–6314.
- CECOTTI, H., SATO-REINHOLD, J., SY, J.L., ELLIOTT, J.C., ECKSTEIN, M.P. & GIESBRECHT, B. (2011b). Impact of target probability on single-trial eeg target detection in a difficult rapid serial visual presentation task. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 6381–6384.
- CECOTTI, H., ECKSTEIN, M.P. & GIESBRECHT, B. (2012). Effects of performing two visual tasks on single-trial detection of event-related potentials. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 1723–1726.
- CECOTTI, H., ECKSTEIN, M.P. & GIESBRECHT, B. (2014a). Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering. *IEEE Transactions on Neural Networks and Learning Systems*, **25**, 2030–2042.
- CECOTTI, H., RIVET, B. *et al.* (2014b). Performance estimation of a cooperative brain-computer interface based on the detection of steady-state visual evoked potentials. In *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, 2078–2082.

REFERENCES

- CECOTTI, H., MARATHE, A. & RIES, A. (2015). Optimization of single-trial detection of event-related potentials through artificial trials. *IEEE Transactions on Biomedical Engineering*, **62**.
- CHEN, X., WANG, Y., NAKANISHI, M., GAO, X., JUNG, T.P. & GAO, S. (2015). High-speed spelling with a noninvasive brain-computer interface. *Proceedings of the National Academy of Sciences*, **112**, E6058–E6067.
- CHOI, H., CHANG, L.H., SHIBATA, K., SASAKI, Y. & WATANABE, T. (2012). Resetting capacity limitations revealed by long-lasting elimination of attentional blink through training. *Proceedings of the National Academy of Sciences*, **109**, 12242–12247.
- CHUN, M. & POTTER, M. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 109–127.
- CICCHETTI, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, **6**, 284.
- CINCOTTI, F., QUITADAMO, L., ALOISE, F., BIANCHI, L., BABILONI, F. & MATTIA, D. (2009). Interacting with the Environment through Non-invasive Brain-Computer Interfaces. In C. Stephanidis, ed., *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*, vol. 5615 of *Lecture Notes in Computer Science*, 483–492, Springer Berlin Heidelberg.
- CINEL, C., POLI, R. & CITI, L. (2004). Possible sources of perceptual errors in P300-based speller paradigm. *Biomedizinische technik*, **49**, 39–40, Proceedings of 2nd International BCI workshop and Training Course.
- CITI, L., POLI, R., CINEL, C. & SEPULVEDA, F. (2008). P300-based BCI mouse with genetically-optimized analogue control. *IEEE transactions on neural systems and rehabilitation engineering*, **16**, 51–61.

REFERENCES

- CITI, L., POLI, R. & CINEL, C. (2009). Exploiting P300 amplitude variations can improve classification accuracy in Donchin’s BCI speller. In *4th International IEEE EMBS Conference on Neural Engineering*, 478–481, Antalya.
- CITI, L., POLI, R. & CINEL, C. (2010). Documenting, modelling and exploiting P300 amplitude changes due to variable target delays in Donchin’s speller. *Journal of Neural Engineering*, **7**, 056006.
- COFFEY, E.B., BROUWER, A.M., WILSCHUT, E.S. & VAN ERP, J.B. (2010). Brain-machine interfaces in space: using spontaneous rather than intentionally generated brain signals. *Acta Astronautica*, **67**, 1–11.
- CONGEDO, M., BARACHANT, A. & ANDREEV, A. (2013). A new generation of brain-computer interface based on riemannian geometry. *arXiv preprint arXiv:1310.8115*.
- CURRAN, E.A. & STOKES, M.J. (2003). Learning to control brain activity: a review of the production and control of EEG components for driving brain-computer interface (BCI) systems. *Brain and cognition*, **51**, 326–336.
- CUTTING, J.E. (2005). Perceiving Scenes in Film and in the World. In J.D. Anderson & B.F. Anderson, eds., *Moving Image Theory: Ecological Considerations*, 9–27, Southern Illinois University Press.
- CUTTING, J.E. (2014). How light and motion bathe the silver screen. *Psychology of Aesthetics, Creativity, and the Arts*, **8**, 340.
- CUTTING, J.E. (2016a). The evolution of pace in popular movies. *Cognitive Research: Principles and Implications*, **1**, 30.
- CUTTING, J.E. (2016b). Narrative theory and the dynamics of popular movies. *Psychonomic Bulletin & Review*, **23**, 1713–1743.
- CUTTING, J.E. & CANDAN, A. (2015). Shot durations, shot classes, and the increased pace of popular movies. *Projections*, **9**, 40–62.
- CUTTING, J.E., BRUNICK, K.L. & DELONG, J.E. (2011). The changing poetics of the dissolve in Hollywood film. *Empirical Studies of the Arts*, **29**, 149–169.

REFERENCES

- CUTTING, J.E., BRUNICK, K.L. & CANDAN, A. (2012). Perceiving event dynamics and parsing Hollywood films. *Journal of Experimental Psychology: Human Perception and Performance*, **38**, 1476.
- DALY, I., NASUTO, S.J. & WARWICK, K. (2012). Brain computer interface control via functional connectivity dynamics. *Pattern Recognition*, **45**, 2123 – 2136, brain Decoding.
- DELL’ACQUA, R., SESSA, P., JOLICŒUR, P. & ROBITAILLE, N. (2006). Spatial attention freezes during the attention blink. *Psychophysiology*, **43**, 394–400.
- DETENBER, B.H., SIMONS, R.F. & BENNETT JR, G.G. (1998). Roll ’em!: The effects of picture motion on emotional responses. *Journal of Broadcasting & Electronic Media*, **42**, 113–127.
- DI LOLLO, V., KAWAHARA, J.I., GHORASHI, S.S. & ENNS, J.T. (2005). The attentional blink: Resource depletion or temporary loss of control? *Psychological research*, **69**, 191–200.
- DIAS, J.C. & PARRA, L.C. (2011). No EEG evidence for subconscious detection during rapid serial visual presentation. In *Signal Processing in Medicine and Biology Symposium (SPMB), 2011 IEEE*, 1–4, IEEE.
- DONCHIN, E. (1981). Surprise!... surprise? *Psychophysiology*, **18**, 493–513.
- DONCHIN, E., RITTER, W. & MCCALLUM, W.C. (1978). The endogenous components of the ERP. *Event-related brain potentials in man*, 349.
- DONCHIN, E., KRAMER, A. & WICKENS, C. (1986). Applications of brain event-related potentials to problems in engineering psychology. *Psychophysiology: Systems, processes, and applications*, 702–718.
- DORR, M., MARTINETZ, T., GEGENFURTNER, K.R. & BARTH, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, **10**, 1–17.
- DYSON, M., SEPULVEDA, F. & GAN, J. (2010). Localisation of cognitive tasks used in eeg-based {BCIs}. *Clinical Neurophysiology*, **121**, 1481 – 1493.

REFERENCES

- ECKSTEIN, M.P., DAS, K., PHAM, B.T., PETERSON, M.F., ABBEY, C.K., SY, J.L. & GIESBRECHT, B. (2012). Neural decoding of collective wisdom with multi-brain computing. *NeuroImage*, **59**, 94–108.
- EIMER, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and clinical neurophysiology*, **99**, 225–234.
- EIMER, M. & KISS, M. (2007). Attentional capture by task-irrelevant fearful faces is revealed by the N2pc component. *Biological psychology*, **74**, 108–112.
- EIMER, M. & KISS, M. (2008). Involuntary attentional capture is determined by task set: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, **20**, 1423–1433.
- EINHÄUSER, W., KOCH, C. & MAKEIG, S. (2007). The duration of the attentional blink in natural scenes depends on stimulus category. *Vision research*, **47**, 597.
- FARNEBÄCK, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis (SCIA 2003)*, 363–370, Springer Berlin Heidelberg, Berlin, Heidelberg.
- FARWELL, L.A. & DONCHIN, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, **70**, 510–523.
- FAZLI, S., GROZEA, C., DANÓCZY, M., BLANKERTZ, B., POPESCU, F. & MÜLLER, K.R. (2009a). Subject independent eeg-based bci decoding. In *Advances in Neural Information Processing Systems*, 513–521.
- FAZLI, S., POPESCU, F., DANÓCZY, M., BLANKERTZ, B., MÜLLER, K.R. & GROZEA, C. (2009b). Subject-independent mental state classification in single trials. *Neural networks*, **22**, 1305–1312.
- FINKE, A., LENHARDT, A. & RITTER, H. (2009). The MindGame: a P300-based brain-computer interface game. *Neural Networks*, **22**, 1329–33.

REFERENCES

- FORSTER, K.I. (1970). Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics*, **8**, 215–221.
- FRANCUZ, P. & ZABIELSKA-MENDYK, E. (2013). Does the Brain Differentiate Between Related and Unrelated Cuts When Processing Audiovisual Messages? An ERP Study. *Media Psychology*, **16**, 461–475.
- GALÁN, F., NUTTIN, M., LEW, E., FERREZ, P.W., VANACKER, G., PHILIPS, J. & MILLÁN, J.D.R. (2008). A brain-actuated wheelchair: asynchronous and non-invasive brain–computer interfaces for continuous control of robots. *Clinical Neurophysiology*, **119**, 2159–2169.
- GERSON, A.D., PARRA, L.C. & SAJDA, P. (2005). Cortical origins of response time variability during rapid discrimination of visual objects. *Neuroimage*, **28**, 342–353.
- GERSON, A.D., PARRA, L.C. & SAJDA, P. (2006). Cortically coupled computer vision for rapid image search. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **14**, 174–179.
- GRUBER, N., MÜRI, R.M., MOSIMANN, U.P., BIERI, R., AESCHIMANN, A., ZITO, G.A., URWYLER, P., NYFFELER, T. & NEF, T. (2014). Effects of age and eccentricity on visual target detection. *Frontiers in aging neuroscience*, **5**, 101.
- HABIB, M., GAYRAUD, D., OLIVA, A., REGIS, J., SALAMON, G. & KHALIL, R. (1991). Effects of handedness and sex on the morphology of the corpus callosum: A study with brain magnetic resonance imaging. *Brain and Cognition*, **16**, 41 – 61.
- HANLEY, J. & MCNEIL, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.
- HASSON, U., NIR, Y., LEVY, I., FUHRMANN, G. & MALACH, R. (2004). Inter-subject Synchronization of Cortical Activity During Natural Vision. *Science*, **303**, 1634–1640.

REFERENCES

- HASSON, U., LANDESMAN, O., KNAPPMAYER, B., VALLINES, I., RUBIN, N. & HEEGER, D.J. (2008). Neurocinematics: The neuroscience of film. *Projections*, **2**, 1–26.
- HEALY, G., WILKINS, P., SMEATON, A.F., IZZO, D., RUCINSKI, M., AMPATZIS, C. & MORAUD, E.M. (2010). Curiosity cloning: Neural modelling for image analysis, technical report. Tech. Rep. 0, Dublin City University; European Space and Technology Research Center (ESTEC), Dublin, Ireland; The Netherlands.
- HICKEY, C., McDONALD, J.J. & THEEUWES, J. (2006). Electrophysiological evidence of the capture of visual attention. *Journal of Cognitive Neuroscience*, **18**, 604–613.
- HILL, J., FARQUHAR, J., MARTENS, S.M., BIESSMANN, F. & SCHÖLKOPF, B. (2008). Effects of stimulus type and of Error-Correcting code design on BCI speller performance. In *The 22nd Annual Conference on Neural Information Processing Systems*, Vancouver.
- HOPE, C., STERR, A., ELANGOVAN, P., GEADES, N., WINDRIDGE, D., YOUNG, K. & WELLS, K. (2013). High throughput screening for mammography using a human-computer interface with rapid serial visual presentation (RSVP). In *SPIE Medical Imaging*, 867303–867303, International Society for Optics and Photonics.
- HOPFINGER, J.B., LUCK, S.J. & HILLYARD, S.A. (2004). Selective Attention: Electrophysiological and Neuromagnetic Studies. In M. Gazzaniga, ed., *The Cognitive Neurosciences III*, chap. 41, The MIT Press.
- HUANG, Y., ERDOGMUS, D., PAVEL, M., MATHAN, S. & HILD II, K.E. (2011). A framework for rapid visual image search using single-trial brain evoked responses. *Neurocomputing*, **74**, 2041–2051.
- HUBER, D.J., KHOSLA, D., MARTIN, K. & CHEN, Y. (2013). A low-bandwidth graphical user interface for high-speed triage of potential items of interest in video imagery. 873608–873608–8.

REFERENCES

- HUGGINS, J. & ZEITLIN, D. (2012). BCI applications. In *Brain-Computer Interfaces: Principles and Practice*, 197–212, Oxford University Press.
- HUGGINS, J.E., MOINUDDIN, A.A., CHIODO, A.E. & WREN, P.A. (2015). What would brain-computer interface users want: Opinions and priorities of potential users with spinal cord injury. *Archives of Physical Medicine and Rehabilitation*, **96**, S38 – S45.e5, the Fifth International Brain-Computer Interface Meeting Presents Clinical and Translational Developments in Brain-Computer Interface Research.
- ISREAL, J.B., CHESNEY, G.L., WICKENS, C.D. & DONCHIN, E. (1980). P300 and Tracking Difficulty: Evidence For Multiple Resources in Dual-Task Performance. *Psychophysiology*, **17**, 259–273.
- ITURRATE, I., CHAVARRIAGA, R., MONTESANO, L., MINGUEZ, J. & MILLÁN, J. (2014). Latency correction of event-related potentials between different experimental protocols. *Journal of neural engineering*, **11**, 036005.
- JIANG, L., WANG, Y., CAI, B., WANG, Y., CHEN, W. & ZHENG, X. (2015). Rapid face recognition based on single-trial event-related potential detection over multiple brains. In *7th International IEEE/EMBS Conference on Neural Engineering (NER)*, 106–109, IEEE.
- JOHNSON, R. (1986). A Triarchic Model of P300 Amplitude. *Psychophysiology*, **23**, 367–384.
- JOLICŒUR, P., SESSA, P., DELL’ACQUA, R. & ROBITAILLE, N. (2006a). Attentional control and capture in the attentional blink paradigm: Evidence from human electrophysiology. *European Journal of Cognitive Psychology*, **18**, 560–578.
- JOLICŒUR, P., SESSA, P., DELL’ACQUA, R. & ROBITAILLE, N. (2006b). On the control of visual spatial attention: Evidence from human electrophysiology. *Psychological research*, **70**, 414–424.

REFERENCES

- JOLICŒUR, P., BRISSON, B. & ROBITAILLE, N. (2008). Dissociation of the N2pc and sustained posterior contralateral negativity in a choice response task. *Brain Research*, **1215**, 160–172.
- JOSHI, A.J., PORIKLI, F. & PAPANIKOLOPOULOS, N.P. (2012). Scalable active learning for multiclass image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**, 2259–2273.
- KAHNEMAN, D., TREISMAN, A. & GIBBS, B.J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive psychology*, **24**, 175–219.
- KALCHER, J. & PFURTSCHELLER, G. (1995). Discrimination between phase-locked and non-phase-locked event-related EEG activity. *Electroencephalography and Clinical Neurophysiology*, **94**, 381–384.
- KANWISHER, N.G. (1987). Repetition blindness: type recognition without token individuation. *Cognition*, **27**, 117–43.
- KAPELLER, C., ORTNER, R., KRAUSZ, G., BRUCKNER, M., ALLISON, B.Z., GUGER, C. & EDLINGER, G. (2014). Toward Multi-brain Communication: Collaborative Spelling with a P300 BCI. In *International Conference on Augmented Cognition*, 47–54.
- KAPOOR, A. & SHENOY, P. (2008). Combining brain computer interfaces with vision for object categorization. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8, IEEE.
- KAPOOR, A., TAN, D., SHENOY, P. & HORVITZ, E. (2008). Complementary computing for visual tasks: Meshing computer vision with human visual processing. *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, 1–7.
- KAPPENMAN, E. & LUCK, S. (2011). ERP components: The ups and downs of brainwave recordings. *Oxford Handbook of ERP Components*, 3–30.

REFERENCES

- KAWAHARA, J.I., KUMADA, T. & DI LOLLO, V. (2006). The attentional blink is governed by a temporary loss of control. *Psychonomic Bulletin & Review*, **13**, 886–890.
- KEATING, P. (2011). The plot point, the darkest moment, and the answered question: three ways of modelling the three-quarter-point. *Journal of screen-writing*, **2**, 85–98.
- KELLY, S.P., LALOR, E.C., FINUCANE, C., MCDARBY, G. & REILLY, R.B. (2005a). Visual spatial attention control in an independent brain-computer interface. *IEEE Transactions on Biomedical Engineering*, **52**, 1588–1596.
- KELLY, S.P., LALOR, E.C., REILLY, R.B. & FOXE, J.J. (2005b). Visual spatial attention tracking using high-density ssvep data for independent brain-computer communication. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **13**, 172–178.
- KLAVER, P., TALSMAN, D., WIJERS, A.A., HEINZE, H.J. & MULDER, G. (1999). An event-related brain potential correlate of visual short-term memory. *NeuroReport*, **10**, 2001–2005.
- KLIMESCH, W. (1996). Memory processes, brain oscillations and EEG synchronization. *International Journal of Psychophysiology*, **24**, 61–100.
- KLIMESCH, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain research reviews*, **29**, 169–195.
- KLIMESCH, W., RUSSEGER, H., DOPPELMAYR, M. & PACHINGER, T. (1998). A method for the calculation of induced band power: implications for the significance of brain oscillations. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, **108**, 123–130.
- KONONENKO, I. & BRATKO, I. (1991). Information-based evaluation criterion for classifier's performance. *Machine Learning*, **6**, 67–80.

REFERENCES

- KORCZOWSKI, L., CONGEDO, M. & JUTTEN, C. (2015). Single-trial classification of multi-user P300-based Brain-Computer Interface using riemannian geometry. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1769–1772.
- KRANCZIOCH, C. & BRYANT, D. (2011). Attentional awakening, resource allocation and the focus of temporal attention. *Neuroreport*, **22**, 161–165.
- KRANCZIOCH, C. & DHINAKARAN, J. (2013). The role of temporal context and expectancy in resource allocation to and perception of rapid serial events. *Brain and Cognition*, **81**, 313–320.
- KRANCZIOCH, C., DEBENER, S., ENGEL, A.K. *et al.* (2003). Event-related potential correlates of the attentional blink phenomenon. *Cognitive Brain Research*, **17**, 177–187.
- KRAULEDAT, M., TANGERMANN, M., BLANKERTZ, B. & MÜLLER, K.R. (2008). Towards zero training for brain-computer interfacing. *PloS one*, **3**, e2967.
- KREPKE, R., BLANKERTZ, B., CURIO, G. & MÜLLER, K.R. (2007). The Berlin Brain-Computer Interface (BBCI) – Towards a new communication channel for online control in gaming applications. *Multimedia Tools and Applications*, **33**, 73–90.
- KRISTJÁNSSON, Á. & CAMPANA, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics*, **72**, 5–18.
- KRISTJÁNSSON, Á. & DRIVER, J. (2008). Priming in visual search: Separating the effects of target repetition, distractor repetition and role-reversal. *Vision research*, **48**, 1217–1232.
- KRISTJÁNSSON, A., EYJÓLFSDÓTTIR, K.O., JÓNSDÓTTIR, A. & ARNKESSON, G. (2010). Temporal consistency is currency in shifts of transient visual attention. *PLoS ONE*, **5**, 1–9.

REFERENCES

- KRUGMAN, H.E. (1971). Brain wave measures of media involvement. *Journal of Advertising Research*, **11**, 3–9.
- KRUSE, A. & MAKEIG, S. (2007). Phase I Analysis Report for UCSD / SoCal NIA Team Acknowledgments. Tech. Rep. January, Institute for Neural Computation, University of California San Diego, La Jolla.
- KUBAT, M. & MATWIN, S. (1997). Addressing the curse of imbalanced data sets: One sided sampling. In *Proceedings of the International Conference on Machine Learning*.
- KUTAS, M. & FEDERMEIER, K.D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, **62**, 621–647.
- LALOR, E.C., KELLY, S.P., FINUCANE, C., BURKE, R., SMITH, R., REILLY, R.B. & MCDARBY, G. (2005). Steady-state vep-based brain-computer interface control in an immersive 3d gaming environment. *EURASIP journal on applied signal processing*, **2005**, 3156–3164.
- LANG, A., GEIGER, S., STRICKWERDA, M. & SUMNER, J. (1993). The effects of related and unrelated cuts on television viewers’ attention, processing capacity, and memory. *Communication Research*, **20**, 4–29.
- LANG, A., BOLLS, P., POTTER, R.F. & KAWAHARA, K. (1999). The effects of production pacing and arousing content on the information processing of television messages. *Journal of Broadcasting & Electronic Media*, **43**, 451–475.
- LANG, A., ZHOU, S., SCHWARTZ, N., BOLLS, P.D. & POTTER, R.F. (2000). The effects of edits on arousal, attention, and memory for television messages: When an edit is an edit can an edit be too much? *Journal of Broadcasting & Electronic Media*, **44**, 94–109.
- LAU, E.F., PHILLIPS, C. & POEPPPEL, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, **9**, 920–933.

REFERENCES

- LAWRENCE, D.H. (1971). Two studies of visual search for word targets with controlled rates of presentation. *Perception & Psychophysics*, **10**, 85–89.
- LI, Y. & NAM, C.S. (2016). Collaborative brain-computer interface for people with motor disabilities. *IEEE Computational Intelligence Magazine*, **11**, 56–66.
- LIN, L., HEDAYAT, A., SINHA, B. & YANG, M. (2002). Statistical methods in assessing agreement: Models, issues, and tools. *Journal of the American Statistical Association*, **97**, 257–270.
- LIN, L.I.K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, **45**, 255–268.
- LIYANAGE, S.R., GUAN, C., ZHANG, H., ANG, K.K., XU, J. & LEE, T.H. (2013). Dynamically weighted ensemble classification for non-stationary EEG processing. *Journal of Neural Engineering*, **10**, 036007.
- LLERA, A., GOMEZ, V. & KAPPEN, H.J. (2012). Adaptive classification on brain-computer interfaces using reinforcement signals. *Neural Computation*, **24**, 2900–2923.
- LUCK, S. (2012). Electrophysiological correlates of the focusing of attention within complex visual scenes: N2pc and related ERP components. *Oxford Handbook of ERP components*.
- LUCK, S.J. (2005). *An introduction to the event-related potential technique*. MIT Press, Cambridge, Massachusetts.
- LUCK, S.J. (2014). *An introduction to the event-related potential technique*. MIT press.
- LUCK, S.J. & HILLYARD, S.A. (1994a). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, **31**, 291–308.
- LUCK, S.J. & HILLYARD, S.A. (1994b). Spatial filtering during visual search: evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 1000–1014.

REFERENCES

- LUCK, S.J., GIRELLI, M., McDERMOTT, M.T. & FORD, M.A. (1997). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive psychology*, **33**, 64–87.
- LYLE, K.B., McCABE, D.P. & ROEDIGER III, H.L. (2008). Handedness is related to memory via hemispheric interaction: evidence from paired associate recall and source memory tasks. *Neuropsychology*, **22**, 523.
- MALJKOVIC, V. & NAKAYAMA, K. (1994). Priming of pop-out: I. role of features. *Memory & Cognition*, **22**, 657–672.
- MALJKOVIC, V. & NAKAYAMA, K. (1996). Priming of pop-out: II. the role of position. *Perception & Psychophysics*, **58**, 977–991.
- MANOR, R. & GEVA, A.B. (2015). Convolutional neural network for multi-category rapid serial visual presentation bci. *Frontiers in computational neuroscience*, **9**.
- MANOR, R., MISHALI, L. & GEVA, A. (2016). Multimodal neural network for rapid serial visual presentation brain computer interface. *Frontiers in Computational Neuroscience*, **10**, 130.
- MARATHE, A.R., LANCE, B.J., McDOWELL, K., NOTHWANG, W.D. & METCALFE, J.S. (2014). Confidence metrics improve human-autonomy integration. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 240–241, ACM.
- MARATHE, A.R., RIES, A.J., LAWHERN, V.J., LANCE, B.J., TOURYAN, J., McDOWELL, K. & CECOTTI, H. (2015). The effect of target and non-target similarity on neural classification performance: A boost from confidence. *Frontiers in Neuroscience*, **9**.
- MARATHE, A.R., LAWHERN, V.J., WU, D., SLAYBACK, D. & LANCE, B.J. (2016). Improved neural signal classification in a rapid serial visual presentation task using active learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **24**, 333–343.

REFERENCES

- MARTENS, S.M.M., HILL, N.J., FARQUHAR, J. & SCHÖLKOPF, B. (2009). Overlap and refractory effects in a brain-computer interface speller based on the visual P300 event-related potential. *Journal of neural engineering*, **6**, 026003.
- MATHAN, S., WHITLOW, S. & ERDOGMUS, D. (2006). Neurophysiologically driven image triage: a pilot study. In *CHI 2006*, 1–6, Montreal, Quebec, Canada.
- MATHAN, S., WHITLOW, S., DORNEICH, M., VERVERS, P. & DAVIS, G. (2007). Neurophysiological estimation of interruptibility: Demonstrating feasibility in a field context. In *In Proceedings of the 4th International Conference of the Augmented Cognition Society*.
- MATHAN, S., ERDOGMUS, D., HUANG, Y., PAVEL, M., VERVERS, P., CARCIOFINI, J., DORNEICH, M. & WHITLOW, S. (2008). Rapid image analysis using neural signals. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, 3309–3314, ACM.
- MCCARTHY, G. & DONCHIN, E. (1981). A metric for thought: a comparison of P300 latency and reaction time. *Science (New York, N.Y.)*, **211**, 77–80.
- MCCOLLOUGH, A.W., MACHIZAWA, M.G. & VOGEL, E.K. (2007). Electrophysiological measures of maintaining representations in visual working memory. *Cortex*, **43**, 77–94.
- MITAL, P.K., SMITH, T.J., HILL, R.L. & HENDERSON, J.M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, **3**, 5–24.
- MORA, N., DE MUNARI, I. & CIAMPOLINI, P. (2015). Subject-independent, ssvep-based bci: trading off among accuracy, responsiveness and complexity. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, 146–149, IEEE.
- MORGAN, S., HANSEN, J. & HILLYARD, S. (1996). Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences*, **93**, 4770–4774.

REFERENCES

- MOST, S.B., SMITH, S.D., COOTER, A.B., LEVY, B.N. & ZALD, D.H. (2007). The naked truth: Positive, arousing distractors impair rapid target perception. *Cognition and Emotion*, **21**, 964–981.
- MURCH, W. (2001). *In the blink of an eye: A perspective on film editing*. Silman-James Press.
- MYRDEN, A. & CHAU, T. (2016). Towards psychologically adaptive brain–computer interfaces. *Journal of Neural Engineering*, **13**, 066022.
- NÄÄTÄNEN, R. (1992). *Attention and brain function*. Psychology Press.
- NAKO, R., WU, R., SMITH, T.J. & EIMER, M. (2014). Item and category-based attentional control during search for real-world objects: Can you find the pants among the pans? *Journal of Experimental Psychology: Human Perception and Performance*, **40**, 1283–1288.
- NAKO, R., SMITH, T.J. & EIMER, M. (2015). Activation of new attentional templates for real-world objects in visual search. *Journal of Cognitive Neuroscience*, **27**, 902–912.
- NEIDER, M.B., ANG, C.W., VOSS, M.W., CARBONARI, R. & KRAMER, A.F. (2013). Training and transfer of training in rapid visual search for camouflaged targets. *PloS one*, **8**, e83885.
- NG, C.A. & LEONG, W. (2014). An EEG-based approach for left-handedness detection. *Biomedical Signal Processing and Control*, **10**, 92 – 101.
- NIELSEN, T., ABEL, A., LORRAIN, D. & MONTPLAISIR, J. (1990). Interhemispheric EEG coherence during sleep and wakefulness in left-and right-handed subjects. *Brain and Cognition*, **14**, 113–125.
- NIJHOLT, A. (2015). Competing and Collaborating Brains: Multi-brain Computer Interfacing. In A.E. Hassanien & A.T. Azar, eds., *Brain-Computer Interfaces*, vol. 74 of *Intelligent Systems Reference Library*, 313–335, Springer International Publishing.

REFERENCES

- NIJHOLT, A. & GÜRKÖK, H. (2013). Multi-brain games: cooperation and competition. In *International Conference on Universal Access in Human-Computer Interaction*, 652–661, Springer.
- NITTONO, H. (2008). Electrophysiology of Kansei: Recent Advances in Event-Related Brain Potential Research. In *Proceedings of the Second International Workshop on Kansei*, 15–18, Fukuoka, Japan.
- PARRA, L., CHRISTOFOROU, C., GERSON, A.D., DYRHOLM, M., LUO, A., WAGNER, M., PHILIASTIDES, M. & SAJDA, P. (2008). Spatiotemporal linear decoding of brain state. *IEEE Signal Processing Magazine*, 107–115.
- PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M. & DUCHESNAY, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, **12**, 2825–2830.
- PENNY, W.D., ROBERTS, S.J., CURRAN, E.A., STOKES, M.J. *et al.* (2000). EEG-based communication: a pattern recognition approach. *IEEE Transactions on Rehabilitation Engineering*, **8**, 214–215.
- PETRICK, M.S. (1981). *Acoustic and semantic encoding during rapid reading*. Ph.D. thesis, Massachusetts Institute of Technology, Department of Psychology.
- PFURTSCHELLER, G. & LOPES DA SILVA, F.H. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, **110**, 1842–1857.
- POGGIO, T. & CAUWENBERGHS, G. (2001). Incremental and decremental support vector machine learning. *Advances in neural information processing systems*, **13**, 409.

REFERENCES

- POHLMAYER, E.A., JANGRAW, D.C., WANG, J., CHANG, S.F. & SAJDA, P. (2010). Combining computer and human vision into a bci: Can the whole be greater than the sum of its parts? In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, 138–141.
- POHLMAYER, E.A., WANG, J., JANGRAW, D.C., LOU, B., CHANG, S.F. & SAJDA, P. (2011). Closing the loop in cortically-coupled computer vision: a brain-computer interface for searching image databases. *Journal of Neural Engineering*, **8**, 036025.
- POLI, R., KENNEDY, J. & BLACKWELL, T. (2007). Particle swarm optimization. *Swarm intelligence*, **1**, 33–57.
- POLI, R., CINEL, C., MATRAN-FERNANDEZ, A., SEPULVEDA, F. & STOICA, A. (2013a). Towards Cooperative Brain-Computer Interfaces for Space Navigation. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI '13)*, 149–160, ACM, Santa Monica, CA, USA.
- POLI, R., CINEL, C., SEPULVEDA, F. & STOICA, A. (2013b). Improving Decision-making based on Visual Perception via a Collaborative Brain-Computer Interface. In *IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, IEEE, San Diego (CA).
- POLI, R., VALERIANI, D. & CINEL, C. (2014). Collaborative brain-computer interface for aiding decision-making. *PloS one*, **9**, e102693.
- POLICH, J. (2004a). Neuropsychology of P3a and P3b: a theoretical overview. *Brainwaves and mind: Recent developments*, 15–29.
- POLICH, J. (2004b). Neuropsychology of P3a and P3b: A theoretical overview. In N.C. Moore & K. Arikan, eds., *Brainwaves and mind: recent developments*, 15–29, Kjellberg Inc.
- POLICH, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, **118**, 2128–2148.

REFERENCES

- POLICH, J. (2011). Neuropsychology of P300. *Oxford Handbook of ERP Components*, 159–188.
- POLICH, J. & CONROY, M. (2003). P3a and P3b from visual stimuli: Gender Effects and normative variability. *The Cognitive Neuroscience of Individual differences*, 293–306.
- POOLMAN, P., FRANK, R.M., LUU, P., PEDERSON, S.M. & TUCKER, D.M. (2008). A single-trial analytic framework for eeg analysis and its application to target detection and classification. *Neuroimage*, **42**, 787–798.
- POTTER, M.C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. In *New methods in reading comprehension Research*, chap. 5, 91–118.
- POTTER, M.C. & LEVY, E. (1969). Recognition memory for a rapid sequence of pictures. *Journal of experimental psychology*, **81**, 10–15.
- PUTZE, F., HILD, J., KÄRGEL, R., HERFF, C., REDMANN, A., BEYERER, J. & SCHULTZ, T. (2013). Locating user attention using eye tracking and EEG for spatio-temporal event selection. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces - IUI '13*, 129–135, ACM Press, Santa Monica, California, USA.
- QUILTER, P., MACGILLIVRAY, B. & WADBROOK, D. (1977). The removal of eye movement artefact from EEG signals using correlation techniques. In *Random Signal Analysis, IEEE Conference Publication*, vol. 159, 93–100.
- RAPPELSBERGER, P. & PETSCHKE, H. (1988). Probability mapping: power and coherence analyses of cognitive processes. *Brain Topography*, **1**, 46–54.
- REEVES, B. & THORSON, E. (1986). WATCHING TELEVISION Experiments on the Viewing Process. *Communication Research*, **13**, 343–361.
- REEVES, B., THORSON, E., ROTHCHILD, M.L., McDONALD, D., HIRSCH, J. & GOLDSTEIN, R. (1985). Attention to television: Intrastimulus effects of movement and scene changes on alpha variation over time. *International Journal of Neuroscience*, **27**, 241–255.

REFERENCES

- REID, V.M. & STRIANO, T. (2008). N400 involvement in the processing of action sequences. *Neuroscience letters*, **433**, 93–97.
- REZA, F.R., Z, A.B., CHRISTOPH, G., W, S.E., C, K.S. & ANDREA, K. (2012). P300 brain computer interface: current challenges and emerging trends. *Frontiers in Neuroengineering*, **5**.
- RICCIO, A., MATTIA, D., SIMIONE, L., OLIVETTI, M. & CINCOTTI, F. (2012). Eye-gaze independent eeg-based brain-computer interfaces for communication. *Journal of Neural Engineering*, **9**, 045001.
- RICCIO, A., SIMIONE, L., SCHETTINI, F., PIZZIMENTI, A., INGHILLERI, M., OLIVETTI BELARDINELLI, M., MATTIA, D. & CINCOTTI, F. (2013). Attention and P300-based BCI performance in people with amyotrophic lateral sclerosis. *Frontiers in Human Neuroscience*, **7**, 732–1–732–9.
- RICCIO, A., PICHIORRI, F., SCHETTINI, F., TOPPI, J., RISETTI, M., FORMISANO, R., MOLINARI, M., ASTOLFI, L., CINCOTTI, F. & MATTIA, D. (2016). Interfacing brain with computer to improve communication and rehabilitation after brain damage. In D. Coyle, ed., *Brain-Computer Interfaces: Lab Experiments to Real-World Applications*, vol. 228 of *Progress in Brain Research*, 357 – 387, Elsevier.
- ROTHSCHILD, M.L., THORSON, E., REEVES, B., HIRSCH, J.E. & GOLDSTEIN, R. (1986). EEG activity and the processing of television commercials. *Communication Research*, **13**, 182–220.
- RUGG, M.D. & COLES, M.G. (1995). *Electrophysiology of mind: Event-related brain potentials and cognition*. Oxford University Press.
- SAJDA, P., POHLMAYER, E., PARRA, L., CHRISTOFOROU, C., DMOCHOWSKI, J., HANNA, B., BAHLMANN, C. & SINGH, M. (2010). In a Blink of an Eye and a Switch of a Transistor: Cortically Coupled Computer Vision. *Proceedings of the IEEE*, **98**, 462–478.

REFERENCES

- SALVARIS, M., CINEL, C., CITI, L. & POLI, R. (2012). Novel protocols for P300-based brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 8–17.
- SAUSENG, P., KLIMESCH, W., SCHABUS, M. & DOPPELMAYR, M. (2005). Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *International Journal of Psychophysiology*, **57**, 97–103.
- SCHACK, B., KLIMESCH, W. & SAUSENG, P. (2005). Phase synchronization between theta and upper alpha oscillations in a working memory task. *International Journal of Psychophysiology*, **57**, 105–114.
- SCHERER, R. & MULLER, G. (2004). An asynchronously controlled EEG-based virtual keyboard: improvement of the spelling rate. *IEEE Transactions on Biomedical Engineering*, **51**, 979–984.
- SCHLÖGL, A., KRONEGG, J., HUGGINS, J.E. & MASON, S.G. (2007). Evaluation criteria for BCI research. In G. Dornhege, J. del R. Millán, T. Hinterberger, D.J. McFarland & K.R. Müller, eds., *Toward brain-computer interfacing*, Neural Information Processing Series, chap. 19, MIT Press.
- SHAPIRO, K.L., RAYMOND, J.E. & ARNELL, K.M. (1994). Attention to visual pattern information produces the attentional blink in rapid serial visual presentation. *Journal of Experimental psychology. Human perception and performance*, **20**, 357–371.
- SHISHKIN, S.L., NUZH DIN, Y.O., SVIRIN, E.P., TROFIMOV, A.G., FEDOROVA, A.A., KOZYRSKIY, B.L. & VELICHKOVSKY, B.M. (2016). EEG Negativity in Fixations Used for Gaze-Based Control: Toward Converting Intentions into Actions with an Eye-Brain-Computer Interface. *Frontiers in Neuroscience*, **10**, 528.
- SIMONS, R.F., DETENBER, B.H., ROEDEMA, T.M. & REISS, J.E. (1999). Emotion processing in three systems: The medium and the message. *Psychophysiology*, **36**, 619–627.

REFERENCES

- SITNIKOVA, T., KUPERBERG, G. & HOLCOMB, P.J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*, **40**, 160–164.
- SITNIKOVA, T., HOLCOMB, P.J., KIYONAGA, K.A. & KUPERBERG, G.R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, **20**, 2037–2057.
- SMITH, M.E. & GEVINS, A. (2004). Attention and brain activity while watching television: Components of viewer engagement. *Media Psychology*, **6**, 285–305.
- SMITH, T. & HENDERSON, J. (2008a). Attentional synchrony in static and dynamic scenes. *Journal of Vision*, **8**, 773–773.
- SMITH, T.J. (2012). The Attentional Theory of Cinematic Continuity. *Projections*, **6**, 1–27.
- SMITH, T.J. (2013). Watching you watch movies: Using eye tracking to inform film theory.
- SMITH, T.J. & HENDERSON, J.M. (2008b). Edit Blindness: The relationship between attention and global change blindness in dynamic scenes. *Journal of Eye Movement Research*, **2**, 1–17.
- SMITH, T.J. & MITAL, P.K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, **13**, 16.
- SMITH, T.J. & YVONNE MARTIN-PORTUGUES SANTACREU, J. (2016). Match-action: the role of motion and audio in creating global change blindness in film. *Media Psychology*, 1–32.
- SMITH, T.J., LEVIN, D. & CUTTING, J.E. (2012). A Window on Reality: Perceiving Edited Moving Images. *Current Directions in Psychological Science*, **21**, 101–106.

REFERENCES

- SOLEYMANI, M., CHANEL, G., KIERKELS, J.J.M. & PUN, T. (2008). Affective characterization of movie scenes based on multimedia content analysis and user's physiological emotional responses. In *2008 10th IEEE International Symposium on Multimedia*, 228–235.
- SUTTER, E.E. (1992). The brain response interface: communication through visually-induced electrical brain responses. *Journal of Microcomputer Applications*, **15**, 31 – 45.
- SWETS, J.A. *et al.* (1988). Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.
- TANG, M.F., BADCOCK, D.R. & VISSER, T.A.W. (2014). Training and the attentional blink: Limits overcome or expectations raised? *Psychonomic Bulletin & Review*, **21**, 406–411.
- TARVAINEN, J., WESTMAN, S. & OITTINEN, P. (2015). The way films feel: Aesthetic features and mood in film. *Psychology of Aesthetics, Creativity, and the Arts*, **9**, 254.
- TEPLAN, M. (2002). Fundamentals of EEG measurement. *Measurement science review*, **2**, 1–11.
- THIERY, T., LAJNEF, T., JERBI, K., ARGUIN, M., AUBIN, M. & JOLICOEUR, P. (2016). Decoding the locus of covert visuospatial attention from eeg signals. *PloS one*, **11**, e0160304.
- THOMPSON, D.E., BLAIN-MORAES, S. & HUGGINS, J.E. (2013). Performance assessment in brain-computer interface-based augmentative and alternative communication. *BioMedical Engineering OnLine*, **12**, 1–23.
- THOMPSON, D.E., QUITADAMO, L.R., MAINARDI, L., UR REHMAN LAGHARI, K., GAO, S., KINDERMANS, P.J., SIMERAL, J.D., FAZEL-REZAI, R., MATTEUCCI, M., FALK, T.H., BIANCHI, L., CHESTEK, C.A. & HUGGINS, J.E. (2014). Performance measurement for brain–computer or brain–machine interfaces: a tutorial. *Journal of Neural Engineering*, **11**, 035001.

REFERENCES

- TIELEMAN, T. & HINTON, G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*, **4**.
- TÖLLNER, T., CONCI, M., RUSCH, T. & MÜLLER, H.J. (2013). Selective manipulation of target identification demands in visual search: the role of stimulus contrast in CDA activations. *Journal of vision*, **13**, 23–23.
- TONG, S. & CHANG, E. (2001). Support vector machine active learning for image retrieval. In *Proceedings of the 9th ACM International Conference on Multimedia*, 107–118, ACM.
- TOURYAN, J., RIES, A.J., WEBER, P. & GIBSON, L. (2013). Integration of automated neural processing into an army-relevant multitasking simulation environment. In *International Conference on Augmented Cognition*, 774–782, Springer.
- TOURYAN, J., APKER, G., LANCE, B.J., KERICK, S.E., RIES, A.J. & MCDOWELL, K. (2014). Estimating endogenous changes in task performance from EEG. *Frontiers in Neuroscience*, **8**.
- TREDER, M.S., SCHMIDT, N.M. & BLANKERTZ, B. (2011). Gaze-independent brain–computer interfaces based on covert attention and feature attention. *Journal of neural engineering*, **8**, 066003.
- TREISMAN, A.M. (1969). Strategies and models of selective attention. *Psychological review*, **76**, 282.
- TROSCIANKO, T., MEESE, T.S. & HINDE, S. (2012). Perception while watching movies: Effects of physical screen size and scene type. *i-Perception*, **3**, 414–425.
- TRUMBO, M.C., MATZEN, L.E., SILVA, A., HAASS, M.J., DIVIS, K. & SPEED, A. (2015). Through a Scanner Quickly: Elicitation of P3 in Transportation Security Officers Following Rapid Image Presentation and Categorization. In *International Conference on Augmented Cognition*, 348–360, Springer.

REFERENCES

- TUCKER, D., ROTH, D. & BAIR, T. (1986). Functional connections among cortical regions: topography of eeg coherence. *Electroencephalography and clinical neurophysiology*, **63**, 242–250.
- UŠĆUMLIĆ, M., CHAVARRIAGA, R. & MILLÁN, J.D.R. (2013). An Iterative Framework for EEG-based Image Search: Robust Retrieval with Weak Classifiers. *PloS one*, **8**, e72018.
- VALERIANI, D., POLI, R. & CINEL, C. (2015a). A collaborative Brain-Computer Interface for improving group detection of visual targets in complex natural environments. In *7th International IEEE/EMBS Conference on Neural Engineering (NER)*, 25–28.
- VALERIANI, D., POLI, R. & CINEL, C. (2015b). A collaborative Brain-Computer Interface to improve human performance in a visual search task. In *7th International IEEE/EMBS Conference on Neural Engineering (NER)*, 218–223.
- VAN ERP, J.B., LOTTE, F. & TANGERMANN, M. (2012). Brain-computer interfaces: beyond medical applications. *Computer-IEEE Computer Society*, **45**, 26–34.
- VIALATTE, F.B., MAURICE, M., DAUWELS, J. & CICHOCKI, A. (2010). Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. *Progress in Neurobiology*, **90**, 418 – 438.
- VIDAL, J.J. (1973). Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, **2**, 157–180.
- VIDAURRE, C., SCHLOGL, A., CABEZA, R., SCHERER, R. & PFURTSCHELLER, G. (2007). Study of on-line adaptive discriminant analysis for EEG-based brain computer interfaces. *IEEE Transactions on Biomedical Engineering*, **54**, 550–556.
- WANG, H.L. & CHEONG, L.F. (2006). Affective understanding in film. *IEEE Transactions on Circuits and Systems for Video Technology*, **16**, 689–704.

REFERENCES

- WANG, Y. & JUNG, T.P. (2011). A Collaborative Brain-Computer Interface for Improving Human Performance. *PLoS ONE*, **6**.
- WANG, Y., WANG, R., GAO, X., HONG, B. & GAO, S. (2006). A practical VEP-based brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **14**, 234–240.
- WEI, Q., HUANG, Y., LI, M. & LU, Z. (2016). VEP-based brain-computer interfaces modulated by Golay complementary series for improving performance. *Technology and Health Care*, 1–9.
- WEISS, S. & RAPPELSBERGER, P. (1996). EEG coherence within the 13–18 Hz band as a correlate of a distinct lexical organisation of concrete and abstract nouns in humans. *Neuroscience letters*, **209**, 17–20.
- WITELSON, S.F. (1985). The brain connection: the corpus callosum is larger in left-handers. *Science*, **229**, 665–668.
- WOLDORFF, M.G. (1993). Distortion of erp averages due to overlap from temporally adjacent erps: analysis and correction. *Psychophysiology*, **30**, 98–119.
- WOLPAW, J., BIRBAUMER, N., HEETDERKS, W., MCFARLAND, D., PECKHAM, P., SCHALK, G., DONCHIN, E., QUATRANO, L., ROBINSON, C. & VAUGHAN, T. (2000). Brain-computer interface technology: a review of the first international meeting. *Rehabilitation Engineering, IEEE Transactions on*, **8**, 164–173.
- WOLPAW, J.R. & WOLPAW, E.W. (2012). Brain-computer interfaces: something new under the sun. *Brain-computer interfaces: principles and practice*, 3–12.
- WOLPAW, J.R., RAMOSER, H., MCFARLAND, D.J. & PFURTSCHELLER, G. (1998). EEG-based communication: improved accuracy by response verification. *IEEE Transactions on Rehabilitation Engineering*, **6**, 326–333.
- WOODMAN, G.F. & LUCK, S.J. (2003). Dissociations among attention, perception, and awareness during object-substitution masking. *Psychological Science*, **14**, 605–611.

REFERENCES

- YASHAR, A. & LAMY, D. (2010a). Intertrial repetition affects perception: The role of focused attention. *Journal of vision*, **10**, 3–3.
- YASHAR, A. & LAMY, D. (2010b). Intertrial repetition facilitates selection in time: Common mechanisms underlie spatial and temporal search. *Psychological Science*, **21**, 243–251.
- YAZDANI, A., VESIN, J.M., IZZO, D., AMPATZIS, C. & EBRAHIMI, T. (2010). Implicit retrieval of salient images using brain computer interface. In *2010 IEEE International Conference on Image Processing*, 3169–3172, IEEE.
- YUAN, P., WANG, Y., WU, W., XU, H., GAO, X. & GAO, S. (2012). Study on an online collaborative BCI to accelerate response to visual targets. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society*, 1736–1739.
- YUAN, P., GAO, X., ALLISON, B., WANG, Y., BIN, G. & GAO, S. (2013). A study of the existing problems of estimating the information transfer rate in online brain–computer interfaces. *Journal of Neural Engineering*, **10**, 026014.
- ZACKS, J.M., SPEER, N., SWALLOW, K. & MALEY, C. (2010). The Brain’s Cutting-Room Floor: Segmentation of Narrative Cinema. *Frontiers in Human Neuroscience*, **4**, 168.
- ZANDER, T.O. & KOTHE, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of Neural Engineering*, **8**, 025005.
- ZANDER, T.O., KOTHE, C., JATZEV, S. & GAERTNER, M. (2010). Enhancing human–computer interaction with input from active and passive brain–computer interfaces. In *Brain-computer interfaces*, 181–199, Springer.
- ZHANG, D., ZHOU, X. & MARTENS, S. (2009). The impact of negative attentional set upon target processing in rsvp: An erp study. *Neuropsychologia*, **47**, 2604–2614.

Appendix A

List of Publications

A.1 Book Chapters

A. Matran-Fernandez, D. Valeriani and R. Poli, “Towards BCIs out of the lab: The impact of motion artefacts on Brain-Computer Interface performance.” In *Wireless medical systems and algorithms: design and applications*, edited by M. Silveira and P. Salvo, CRC Press.

A.2 Journal Papers

A. Matran-Fernandez and R. Poli, “Brain-Computer Interfaces for Detection and Localisation of Targets in Aerial Images,” *IEEE Transactions on Biomedical Engineering*, vol. XX, no. XX, 2016 (pre-print available online).

A.3 Conference Papers

A. Matran-Fernandez and R. Poli, “Event-Related Potentials induced by Cuts in Feature Movies and their Exploitation for Understanding Cut Efficacy,” *7th International IEEE/EMBS Conference on Neural Engineering (NER)*, Montpellier, France, 2015.

A. Matran-Fernandez and R. Poli, “Collaborative Brain-Computer Interfaces for Target Localisation in Rapid Serial Visual Presentation,” *6th Computer Science and Electronic Engineering Conference (CEEC)*, Colchester, UK, 2014, pp. 127–132.

A. Matran-Fernandez, R. Poli, and C. Cinel, “Collaborative brain-computer interfaces for the automatic classification of images,” *6th International IEEE/EMBS Conference on Neural Engineering (NER)*, San Diego, USA, 2013, pp. 1096–1099.

Appendix B

Comparisons Between the Collaborative BCI Methods

This appendix complements the results from Chapters 3-4, where the different methods of combining evidence to form collaborative BCIs were compared.

The tables included in this appendix report the results of pairwise comparisons between the three modes of creating cBCIs: LDA-cBCI (“L”), MC-cBCI (“M”) and SC-cBCI (“S”). The symbol $>$ indicates that the method on the left is significantly better than the one of the right (at the 5% significance level). These values are also highlighted in bold face. However, if method names are separated by the symbol $=$, they are not significantly different. N/A values are used to indicate cases in which there were not enough samples to perform the test. Numbers in the tables represent p values from one-sided paired Wilcoxon tests comparing pairs of methods (after Bonferroni correction for multiple comparisons).

Comparisons between the three modes of creating cBCIs for target detection. See text for full description.

Method	Size	Difficulty level						
		1	2	3	4	5	6	7
MC-cBCI vs SC-cBCI	2	M=S 0.997	M=S 0.096	M=S 1	M=S 1	M=S 1	M=S 0.105	M=S 1
	3	M>S < 0.001	M>S < 0.001	M>S 0.005	M>S < 0.001	S>M < 0.001	M>S < 0.001	M>S 0.016
	4	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	S>M < 0.001	M>S < 0.001	M>S < 0.001
	5	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	S>M < 0.001	M>S < 0.001	M>S < 0.001
	6	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S 0.065	M>S < 0.001	M>S < 0.001
	7	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M=S 1	M>S < 0.001	M>S < 0.001
	8	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M=S 1	M>S < 0.001	M>S < 0.001
	9	M>S < 0.001	N/A –	M>S < 0.001	N/A –	N/A –	N/A –	N/A –
LDA-cBCI vs SC-cBCI	2	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	S=L 0.339	L>S < 0.001	S=L 1
	3	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S 0.000467
	4	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001
	5	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001
	6	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001
	7	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001
	8	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001
	9	L>S < 0.001	N/A –	L>S < 0.001	N/A –	N/A –	N/A –	N/A –
LDA-cBCI vs MC-cBCI	2	L>M < 0.001	L>M 0.012	L>M 0.025	L>M < 0.001	L>M < 0.001	L>M < 0.001	M=L 1
	3	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	M=L 1
	4	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M 0.009
	5	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001
	6	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001
	7	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001
	8	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M 0.010
	9	L>M < 0.001	N/A –	L>M < 0.001	N/A –	N/A –	N/A –	N/A –

Comparisons between the three modes of creating cBCIs for target localisation.
See text for full description.

Method	Size	Difficulty level						
		1	2	3	4	5	6	7
MC-cBCI vs SC-cBCI	2	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S 0.007
	3	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001
	4	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001
	5	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001
	6	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001
	7	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001
	8	M>S < 0.001	M>S < 0.001	M>S < 0.001	M>S < 0.001	N/A –	M>S < 0.001	M>S < 0.001
	9	M>S < 0.001	N/A –	M>S < 0.001	N/A –	N/A –	N/A –	N/A –
LDA-cBCI vs SC-cBCI	2	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S 0.001
	3	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S 0.375	L>S < 0.001	L>S < 0.001
	4	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S 0.375	L>S < 0.001	L>S < 0.001
	5	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S 0.375	L>S < 0.001	L>S < 0.001
	6	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S 0.375	L>S < 0.001	L>S < 0.001
	7	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S 0.375	L>S < 0.001	L>S < 0.001
	8	L>S < 0.001	L>S < 0.001	L>S < 0.001	L>S < 0.001	N/A –	L>S < 0.001	L>S < 0.001
	9	L>S < 0.001	N/A –	L>S < 0.001	N/A –	N/A –	N/A –	N/A –
LDA-cBCI vs MC-cBCI	2	M=L 1	M=L 1	M=L 1	L>M 0.010	M=L 1	M=L 1	M=L 1
	3	L>M < 0.001	L>M 0.002	M=L 1	L>M < 0.001	L>M < 0.001	L>M < 0.001	M=L 1
	4	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	M=L 1
	5	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001
	6	L>M < 0.001	L>M < 0.001	M=L 1	L>M < 0.001	L>M < 0.001	L>M < 0.001	L>M < 0.001
	7	M=L 0.384	M=L 1	L>M 0.008	L>M < 0.001	L>M < 0.001	M=L 1	M=L 1
	8	M=L 1	M=L 0.055	M>L < 0.001	M=L 1	N/A –	M=L 0.083	L>M 0.002
	9	M=L 1	N/A –	M=L 1	N/A –	N/A –	N/A –	N/A –

Appendix C

Participant Selection results

This appendix complements the results from Chapter 6. The two sections of the appendix focus on the target detection (Section C.1) and the target location systems (Section C.2). For each of them, we start reporting the full results on the effects of group member selection by providing the median AUC values for each difficulty level and group size, for different values of the dissimilarity index together with the percentages of groups (with respect to the maximum groups that can be formed for each group size) that are accepted for each value of δ considered. These subsections are then followed by level-by-level comparisons between the AUCs achieved by the cBCIs and the average AUC performance of the groups. Finally, each section concludes with tables that report the level-by-level improvements that are obtained by the cBCIs with respect to the best performer of the groups.

C.1 Group Member Selection on the Target Detection System

C.1.1 Median AUCs with Group Member Selection

The tables from this section report the median AUC performance for each level and group size, as a function of the dissimilarity threshold, as well as the percentages of groups that are accepted for each level and value of δ .

Median AUC values for the three types of cBCIs for target vs non-target classification for difficulty level 1, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	0.95	0.97	0.98	0.98	–	–	–	–	–	–
	10%	0.95	0.97	0.98	0.99	0.99	0.99	–	–	–	–
	15%	0.94	0.97	0.98	0.98	0.99	0.99	–	–	–	–
	20%	0.93	0.96	0.97	0.98	0.98	0.99	0.99	0.99	–	–
	25%	0.92	0.94	0.96	0.97	0.97	0.98	0.98	0.98	–	–
	100%	0.92	0.94	0.95	0.96	0.97	0.98	0.98	0.98	0.99	0.99
MC-cBCI	5%	0.96	0.97	0.98	0.99	–	–	–	–	–	–
	10%	0.96	0.97	0.98	0.99	0.99	1.00	–	–	–	–
	15%	0.95	0.97	0.98	0.99	0.99	1.00	–	–	–	–
	20%	0.94	0.96	0.97	0.97	0.98	0.98	0.99	0.99	–	–
	25%	0.92	0.94	0.96	0.97	0.97	0.98	0.98	0.98	–	–
	100%	0.92	0.94	0.96	0.96	0.97	0.98	0.98	0.98	0.99	0.99
LDA-cBCI	5%	0.96	0.98	0.99	0.99	–	–	–	–	–	–
	10%	0.96	0.98	0.99	0.99	1.00	1.00	–	–	–	–
	15%	0.96	0.98	0.99	0.99	1.00	1.00	–	–	–	–
	20%	0.94	0.97	0.98	0.99	0.99	0.99	0.99	1.00	–	–
	25%	0.94	0.96	0.97	0.98	0.98	0.99	0.99	0.99	–	–
	100%	0.94	0.96	0.97	0.98	0.98	0.99	0.99	0.99	0.99	0.99

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 1.

δ	Group size									
	2	3	4	5	6	7	8	9	10	11
5%	32%	9%	2%	0%	0%	0%	0%	0%	0%	0%
10%	49%	23%	10%	4%	1%	0%	0%	0%	0%	0%
15%	60%	29%	12%	4%	1%	0%	0%	0%	0%	0%
20%	81%	60%	42%	28%	18%	10%	5%	1%	0%	0%
25%	92%	80%	65%	50%	34%	21%	10%	3%	0%	0%

Median AUC values for the three types of cBCIs for target vs non-target classification for difficulty level 3, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	0.83	0.88	0.90	–	–	–	–	–	–	–
	10%	0.78	0.83	0.88	0.90	–	–	–	–	–	–
	15%	0.78	0.82	0.78	0.76	0.75	–	–	–	–	–
	20%	0.79	0.83	0.85	0.88	0.90	0.91	–	–	–	–
	25%	0.78	0.82	0.84	0.86	0.87	0.89	0.90	0.90	–	–
	100%	0.78	0.82	0.83	0.85	0.86	0.87	0.88	0.88	0.90	0.90
MC-cBCI	5%	0.85	0.90	0.92	–	–	–	–	–	–	–
	10%	0.78	0.85	0.92	0.93	–	–	–	–	–	–
	15%	0.78	0.82	0.78	0.78	0.78	–	–	–	–	–
	20%	0.79	0.84	0.87	0.90	0.92	0.93	–	–	–	–
	25%	0.79	0.83	0.85	0.87	0.89	0.90	0.91	0.91	–	–
	100%	0.79	0.83	0.86	0.87	0.89	0.90	0.91	0.91	0.92	0.92
LDA-cBCI	5%	0.85	0.90	0.92	–	–	–	–	–	–	–
	10%	0.78	0.85	0.92	0.93	–	–	–	–	–	–
	15%	0.78	0.82	0.78	0.79	0.79	–	–	–	–	–
	20%	0.80	0.85	0.88	0.91	0.92	0.93	–	–	–	–
	25%	0.80	0.84	0.87	0.88	0.90	0.91	0.92	0.93	–	–
	100%	0.80	0.84	0.87	0.89	0.90	0.91	0.92	0.92	0.93	0.93

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 3.

δ	Group size									
	2	3	4	5	6	7	8	9	10	11
5%	25%	3%	0%	0%	0%	0%	0%	0%	0%	0%
10%	43%	12%	2%	0%	0%	0%	0%	0%	0%	0%
15%	58%	25%	8%	1%	0%	0%	0%	0%	0%	0%
20%	72%	43%	20%	8%	2%	0%	0%	0%	0%	0%
25%	92%	80%	65%	50%	34%	21%	10%	3%	0%	0%

Median AUC values for the three types of cBCIs for target vs non-target classification for difficulty level 4, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	0.85	0.89	0.90	0.91	—	—	—	—	—
	10%	0.85	0.89	0.90	0.91	0.92	—	—	—	—
	15%	0.84	0.88	0.89	0.91	0.92	—	—	—	—
	20%	0.84	0.88	0.90	0.91	0.92	0.92	—	—	—
	25%	0.82	0.84	0.88	0.90	0.90	0.91	—	—	—
	100%	0.80	0.83	0.86	0.87	0.89	0.90	0.90	0.91	0.93
MC-cBCI	5%	0.87	0.88	0.89	0.90	—	—	—	—	—
	10%	0.87	0.88	0.89	0.90	0.91	—	—	—	—
	15%	0.84	0.87	0.89	0.90	0.91	—	—	—	—
	20%	0.84	0.87	0.89	0.90	0.91	0.92	—	—	—
	25%	0.81	0.86	0.89	0.90	0.91	0.91	—	—	—
	100%	0.80	0.84	0.86	0.88	0.90	0.91	0.91	0.92	0.93
LDA-cBCI	5%	0.87	0.89	0.90	0.91	—	—	—	—	—
	10%	0.87	0.89	0.90	0.91	0.91	—	—	—	—
	15%	0.85	0.89	0.90	0.91	0.91	—	—	—	—
	20%	0.85	0.89	0.90	0.91	0.92	0.92	—	—	—
	25%	0.84	0.88	0.90	0.91	0.91	0.92	—	—	—
	100%	0.83	0.87	0.89	0.90	0.91	0.92	0.92	0.92	0.93

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 4.

δ	Group size								
	2	3	4	5	6	7	8	9	10
5%	35%	13%	4%	0%	0%	0%	0%	0%	0%
10%	42%	17%	7%	2%	0%	0%	0%	0%	0%
15%	55%	26%	11%	4%	0%	0%	0%	0%	0%
20%	60%	32%	17%	8%	3%	0%	0%	0%	0%
25%	77%	52%	31%	16%	6%	1%	0%	0%	0%

Median AUC values for the three types of cBCIs for target vs non-target classification for difficulty level 5, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	0.58	0.54	0.55	—	—	—	—	—	—
	10%	0.59	0.60	0.59	0.60	0.59	0.60	—	—	—
	15%	0.59	0.63	0.62	0.61	0.61	0.61	0.63	—	—
	20%	0.59	0.64	0.64	0.65	0.65	0.65	0.65	0.63	0.63
	25%	0.59	0.64	0.64	0.65	0.65	0.65	0.65	0.63	0.63
	100%	0.59	0.64	0.64	0.65	0.65	0.65	0.65	0.63	0.63
MC-cBCI	5%	0.58	0.57	0.57	—	—	—	—	—	—
	10%	0.58	0.59	0.59	0.59	0.59	0.59	—	—	—
	15%	0.59	0.60	0.62	0.62	0.63	0.63	0.63	—	—
	20%	0.60	0.62	0.63	0.64	0.64	0.65	0.66	0.66	0.67
	25%	0.60	0.62	0.63	0.64	0.64	0.65	0.66	0.66	0.67
	100%	0.60	0.62	0.63	0.64	0.64	0.65	0.66	0.66	0.67
LDA-cBCI	5%	0.59	0.60	0.59	—	—	—	—	—	—
	10%	0.59	0.60	0.60	0.61	0.61	0.62	—	—	—
	15%	0.59	0.61	0.63	0.63	0.64	0.64	0.66	—	—
	20%	0.60	0.65	0.67	0.68	0.68	0.69	0.70	0.70	0.71
	25%	0.60	0.65	0.67	0.68	0.68	0.69	0.70	0.70	0.71
	100%	0.60	0.65	0.67	0.68	0.68	0.69	0.70	0.70	0.71

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 5.

δ	Group size								
	2	3	4	5	6	7	8	9	10
5%	37%	9%	1%	0%	0%	0%	0%	0%	0%
10%	77%	51%	30%	15%	6%	1%	0%	0%	0%
15%	88%	72%	54%	38%	23%	12%	4%	0%	0%
20%	100%	100%	100%	100%	100%	100%	100%	100%	100%
25%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Median AUC values for the three types of cBCIs for target vs non-target classification for difficulty level 6, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	0.72	0.74	0.71	–	–	–	–	–	–
	10%	0.66	0.73	0.78	0.79	–	–	–	–	–
	15%	0.67	0.73	0.78	0.79	0.81	–	–	–	–
	20%	0.66	0.69	0.71	0.71	0.72	0.72	0.73	–	–
	25%	0.66	0.69	0.71	0.71	0.72	0.73	0.71	0.72	–
	100%	0.67	0.71	0.72	0.74	0.75	0.76	0.77	0.79	0.79
MC-cBCI	5%	0.74	0.78	0.80	–	–	–	–	–	–
	10%	0.71	0.76	0.79	0.82	–	–	–	–	–
	15%	0.69	0.74	0.78	0.81	0.83	–	–	–	–
	20%	0.68	0.71	0.73	0.75	0.76	0.78	0.78	–	–
	25%	0.68	0.71	0.73	0.75	0.76	0.77	0.78	0.79	–
	100%	0.69	0.72	0.75	0.77	0.78	0.80	0.81	0.82	0.83
LDA-cBCI	5%	0.74	0.79	0.82	–	–	–	–	–	–
	10%	0.71	0.75	0.80	0.83	–	–	–	–	–
	15%	0.71	0.76	0.80	0.82	0.84	–	–	–	–
	20%	0.69	0.73	0.74	0.76	0.77	0.78	0.79	–	–
	25%	0.69	0.73	0.74	0.76	0.77	0.77	0.78	0.79	–
	100%	0.71	0.74	0.76	0.78	0.80	0.81	0.81	0.82	0.83

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 6.

δ	Group size								
	2	3	4	5	6	7	8	9	10
5%	24%	5%	0%	0%	0%	0%	0%	0%	0%
10%	44%	14%	3%	0%	0%	0%	0%	0%	0%
15%	66%	35%	14%	4%	0%	0%	0%	0%	0%
20%	88%	72%	54%	38%	23%	12%	4%	0%	0%
25%	93%	82%	69%	55%	42%	30%	20%	10%	0%

Median AUC values for the three types of cBCIs for target vs non-target classification for difficulty level 7, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	0.64	0.69	0.74	0.76	–	–	–	–	–
	10%	0.62	0.63	0.64	0.65	0.67	0.66	0.68	0.74	–
	15%	0.61	0.62	0.63	0.64	0.65	0.66	0.66	0.69	0.70
	20%	0.61	0.62	0.63	0.64	0.65	0.66	0.66	0.69	0.70
	25%	0.61	0.62	0.63	0.64	0.65	0.66	0.66	0.69	0.70
	100%	0.61	0.62	0.63	0.64	0.65	0.66	0.66	0.69	0.70
MC-cBCI	5%	0.64	0.68	0.72	0.74	–	–	–	–	–
	10%	0.62	0.64	0.66	0.68	0.69	0.70	0.71	0.73	–
	15%	0.61	0.63	0.65	0.66	0.68	0.70	0.71	0.72	0.72
	20%	0.61	0.63	0.65	0.66	0.68	0.70	0.71	0.72	0.72
	25%	0.61	0.63	0.65	0.66	0.68	0.70	0.71	0.72	0.72
	100%	0.61	0.63	0.65	0.66	0.68	0.70	0.71	0.72	0.72
LDA-cBCI	5%	0.64	0.68	0.74	0.75	–	–	–	–	–
	10%	0.62	0.65	0.67	0.69	0.70	0.71	0.72	0.74	–
	15%	0.61	0.63	0.66	0.68	0.69	0.70	0.72	0.72	0.73
	20%	0.61	0.63	0.66	0.68	0.69	0.70	0.72	0.72	0.73
	25%	0.61	0.63	0.66	0.68	0.69	0.70	0.72	0.72	0.73
	100%	0.61	0.63	0.66	0.68	0.69	0.70	0.72	0.72	0.73

r

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 7.

	Group size								
δ	2	3	4	5	6	7	8	9	10
5%	35%	11%	2%	0%	0%	0%	0%	0%	0%
10%	91%	78%	64%	51%	40%	30%	20%	10%	0%
15%	100%	100%	100%	100%	100%	100%	100%	100%	100%
20%	100%	100%	100%	100%	100%	100%	100%	100%	100%
25%	100%	100%	100%	100%	100%	100%	100%	100%	100%

C.1.2 Comparison with $\text{avg}(AUC_1, AUC_2, \dots, AUC_r)$

This section contains the results for the target detection system. In particular, here we report, for the target detection system, the median improvements (in percentage) when using collaborative BCIs over the average AUC performance of the group, for each level and group size.

Median improvements over the average participant in the group when using collaborative BCIs for target detection at difficulty level 1, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	+1.1%	+2.7%	+3.9%	+4.0%	–	–	–	–	–	–
	10%	+1.2%	+2.7%	+3.6%	+4.1%	+4.4%	+4.8%	–	–	–	–
	15%	+1.6%	+3.1%	+3.7%	+4.1%	+4.4%	+4.8%	–	–	–	–
	20%	+3.6%	+6.0%	+7.2%	+7.8%	+8.4%	+9.2%	+9.4%	+9.6%	–	–
	25%	+4.2%	+7.0%	+9.2%	+10.2%	+11.1%	+11.4%	+12.2%	+13.1%	–	–
	100%	+4.5%	+8.7%	+10.6%	+12.2%	+13.0%	+13.7%	+14.2%	+14.3%	+15.3%	+15.1%
MC-cBCI	5%	+2.1%	+4.3%	+5.5%	+5.7%	–	–	–	–	–	–
	10%	+2.3%	+4.1%	+5.4%	+5.7%	+6.4%	+6.5%	–	–	–	–
	15%	+2.6%	+4.9%	+5.5%	+5.8%	+6.4%	+6.5%	–	–	–	–
	20%	+5.0%	+7.3%	+8.8%	+9.5%	+10.2%	+11.2%	+11.3%	+11.4%	–	–
	25%	+5.7%	+8.2%	+10.4%	+11.8%	+12.7%	+13.0%	+13.8%	+14.5%	–	–
	100%	+5.9%	+9.7%	+11.5%	+13.3%	+14.2%	+15.0%	+15.6%	+15.6%	+16.7%	+16.5%
LDA-cBCI	5%	+2.5%	+4.7%	+5.7%	+5.8%	–	–	–	–	–	–
	10%	+2.8%	+4.7%	+5.4%	+5.8%	+6.4%	+6.6%	–	–	–	–
	15%	+3.0%	+5.0%	+5.6%	+5.9%	+6.4%	+6.6%	–	–	–	–
	20%	+5.6%	+8.4%	+9.5%	+9.9%	+10.3%	+11.3%	+11.3%	+11.4%	–	–
	25%	+6.7%	+10.0%	+11.4%	+12.8%	+13.2%	+13.4%	+14.2%	+15.0%	–	–
	100%	+7.2%	+10.8%	+13.1%	+14.4%	+15.1%	+15.8%	+16.2%	+16.1%	+17.1%	+16.9%

Median improvements over the average participant in the group when using collaborative BCIs for target detection at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	+4.9%	+6.4%	+7.7%	+8.7%	–	–	–	–	–
	10%	+4.4%	+6.3%	+7.2%	+7.8%	+8.9%	–	–	–	–
	15%	+5.2%	+7.1%	+7.9%	+8.7%	+10.2%	–	–	–	–
	20%	+5.4%	+7.6%	+9.3%	+10.2%	+10.3%	+11.0%	–	–	–
	25%	+5.7%	+9.4%	+11.0%	+12.9%	+14.0%	+15.3%	+16.4%	–	–
	100%	+6.6%	+11.0%	+13.2%	+14.7%	+16.1%	+17.0%	+17.8%	+18.7%	+18.6%
MC-cBCI	5%	+6.2%	+9.2%	+10.1%	+10.5%	–	–	–	–	–
	10%	+6.2%	+8.7%	+9.1%	+9.5%	+9.7%	–	–	–	–
	15%	+7.4%	+9.5%	+10.1%	+10.5%	+11.7%	–	–	–	–
	20%	+7.6%	+10.1%	+11.0%	+12.0%	+12.5%	+12.5%	–	–	–
	25%	+8.0%	+11.6%	+13.7%	+15.2%	+16.4%	+17.3%	+17.8%	–	–
	100%	+9.4%	+13.4%	+16.1%	+17.5%	+18.6%	+19.2%	+19.9%	+20.2%	+20.2%
LDA-cBCI	5%	+6.3%	+9.3%	+10.5%	+10.6%	–	–	–	–	–
	10%	+6.5%	+8.8%	+9.0%	+9.7%	+9.8%	–	–	–	–
	15%	+7.6%	+9.8%	+10.5%	+10.6%	+11.8%	–	–	–	–
	20%	+8.2%	+10.6%	+11.2%	+12.3%	+12.8%	+12.5%	–	–	–
	25%	+9.3%	+13.0%	+14.7%	+16.3%	+17.4%	+18.0%	+18.5%	–	–
	100%	+10.7%	+15.0%	+17.7%	+18.8%	+20.1%	+20.4%	+21.0%	+21.2%	+21.0%

Median improvements over the average participant in the group when using collaborative BCIs for target detection at difficulty level 3, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	+5.4%	+9.7%	+11.6%	–	–	–	–	–	–	–
	10%	+3.4%	+8.5%	+11.1%	+12.1%	–	–	–	–	–	–
	15%	+4.6%	+8.2%	+10.4%	+11.6%	+12.9%	–	–	–	–	–
	20%	+4.8%	+8.4%	+11.5%	+13.3%	+14.7%	+14.6%	–	–	–	–
	25%	+5.4%	+9.4%	+12.3%	+14.6%	+16.4%	+17.7%	+18.8%	+19.7%	–	–
	100%	+5.4%	+9.6%	+12.8%	+15.2%	+16.9%	+18.3%	+19.4%	+20.4%	+21.1%	+21.4%
MC-cBCI	5%	+6.6%	+12.0%	+14.5%	–	–	–	–	–	–	–
	10%	+4.3%	+9.7%	+12.8%	+14.9%	–	–	–	–	–	–
	15%	+4.8%	+8.4%	+8.0%	+8.3%	+9.6%	–	–	–	–	–
	20%	+5.1%	+9.4%	+12.3%	+15.0%	+16.6%	+17.8%	–	–	–	–
	25%	+5.4%	+10.0%	+13.0%	+15.4%	+17.2%	+18.7%	+19.8%	+21.0%	–	–
	100%	+5.4%	+9.9%	+12.9%	+15.2%	+16.9%	+18.4%	+19.7%	+20.7%	+21.5%	+22.3%
LDA-cBCI	5%	+7.4%	+12.5%	+14.9%	–	–	–	–	–	–	–
	10%	+5.2%	+10.1%	+13.3%	+15.4%	–	–	–	–	–	–
	15%	+6.5%	+10.1%	+11.2%	+12.2%	+13.5%	–	–	–	–	–
	20%	+7.1%	+11.7%	+14.5%	+16.7%	+18.2%	+19.1%	–	–	–	–
	25%	+7.9%	+13.2%	+16.5%	+18.5%	+20.2%	+21.5%	+22.3%	+23.3%	–	–
	100%	+8.3%	+14.0%	+17.7%	+20.2%	+22.0%	+23.2%	+24.2%	+25.0%	+25.9%	+26.1%

Median improvements over the average participant in the group when using collaborative BCIs for target detection at difficulty level 4, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	+1.4%	+6.0%	+8.3%	+10.3%	–	–	–	–	–
	10%	+1.8%	+5.9%	+8.3%	+10.4%	+11.1%	–	–	–	–
	15%	+2.7%	+6.7%	+9.1%	+10.9%	+12.3%	–	–	–	–
	20%	+2.9%	+7.3%	+9.9%	+12.3%	+13.6%	+14.7%	–	–	–
	25%	+2.9%	+7.3%	+10.2%	+12.6%	+13.7%	+15.7%	–	–	–
	100%	+3.9%	+9.3%	+12.4%	+15.3%	+17.3%	+18.5%	+20.0%	+20.7%	+22.6%
MC-cBCI	5%	+4.3%	+8.8%	+11.2%	+12.7%	–	–	–	–	–
	10%	+4.2%	+8.4%	+11.2%	+12.7%	+13.5%	–	–	–	–
	15%	+4.5%	+8.8%	+11.6%	+13.1%	+14.6%	–	–	–	–
	20%	+4.5%	+9.1%	+11.6%	+13.3%	+15.5%	+15.7%	–	–	–
	25%	+4.3%	+9.1%	+11.9%	+14.8%	+15.8%	+17.0%	–	–	–
	100%	+4.6%	+9.7%	+13.5%	+16.2%	+18.4%	+20.0%	+21.5%	+22.6%	+23.3%
LDA-cBCI	5%	+4.8%	+9.5%	+12.3%	+13.5%	–	–	–	–	–
	10%	+4.8%	+9.2%	+12.0%	+13.4%	+14.0%	–	–	–	–
	15%	+5.8%	+9.8%	+12.5%	+13.7%	+15.5%	–	–	–	–
	20%	+5.8%	+10.1%	+12.7%	+14.0%	+16.5%	+16.8%	–	–	–
	25%	+6.6%	+11.8%	+15.4%	+16.9%	+17.7%	+19.0%	–	–	–
	100%	+7.7%	+15.3%	+18.6%	+21.1%	+23.1%	+24.4%	+25.1%	+26.5%	+26.8%

Median improvements over the average participant in the group when using collaborative BCIs for target detection at difficulty level 5, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	
SC-cBCI	5%	-2.1%	-4.9%	+0.7%	—	—	—	—	—	—
	10%	-0.4%	+3.3%	+6.2%	+5.6%	+6.3%	+10.6%	—	—	—
	15%	-0.4%	+2.9%	+6.0%	+5.9%	+7.4%	+8.9%	+9.9%	—	—
	20%	-0.4%	+4.4%	+6.9%	+8.7%	+10.7%	+12.1%	+12.7%	+13.6%	+10.6%
	25%	-0.4%	+4.4%	+6.9%	+8.7%	+10.7%	+12.1%	+12.7%	+13.6%	+10.6%
	100%	-0.4%	+4.4%	+6.9%	+8.7%	+10.7%	+12.1%	+12.7%	+13.6%	+10.6%
MC-cBCI	5%	+0.6%	-2.6%	-2.7%	—	—	—	—	—	—
	10%	+0.6%	+2.6%	+2.8%	+3.9%	+4.9%	+5.4%	—	—	—
	15%	+0.6%	+2.6%	+4.2%	+4.8%	+6.5%	+7.4%	+8.6%	—	—
	20%	+0.6%	+2.9%	+5.2%	+7.1%	+8.7%	+10.3%	+11.5%	+12.9%	+14.1%
	25%	+0.6%	+2.9%	+5.2%	+7.1%	+8.7%	+10.3%	+11.5%	+12.9%	+14.1%
	100%	+0.6%	+2.9%	+5.2%	+7.1%	+8.7%	+10.3%	+11.5%	+12.9%	+14.1%
LDA-cBCI	5%	+1.3%	+3.0%	+2.8%	—	—	—	—	—	—
	10%	+2.5%	+5.5%	+7.6%	+8.5%	+9.9%	+11.1%	—	—	—
	15%	+3.1%	+6.4%	+8.6%	+10.5%	+11.6%	+13.0%	+13.8%	—	—
	20%	+3.6%	+7.8%	+10.5%	+12.8%	+15.3%	+17.0%	+18.4%	+19.7%	+20.5%
	25%	+3.6%	+7.8%	+10.5%	+12.8%	+15.3%	+17.0%	+18.4%	+19.7%	+20.5%
	100%	+3.6%	+7.8%	+10.5%	+12.8%	+15.3%	+17.0%	+18.4%	+19.7%	+20.5%

Median improvements over the average participant in the group when using collaborative BCIs for target detection at difficulty level 6, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	+3.3%	+9.5%	+15.8%	–	–	–	–	–	–
	10%	+7.2%	+11.2%	+15.5%	+17.1%	–	–	–	–	–
	15%	<i>+7.8%</i>	+13.6%	+16.8%	+19.3%	+20.9%	–	–	–	–
	20%	+7.7%	+11.8%	+15.4%	+18.2%	+19.9%	+20.9%	+24.0%	–	–
	25%	+7.8%	+12.7%	+16.3%	+18.7%	+20.6%	+22.1%	+23.2%	+24.4%	–
	100%	+8.0%	+13.4%	+16.7%	+19.4%	+21.5%	+22.9%	+25.3%	+27.2%	+27.7%
MC-cBCI	5%	<i>+5.6%</i>	+10.5%	+14.7%	–	–	–	–	–	–
	10%	+8.5%	+12.2%	+16.2%	+18.9%	–	–	–	–	–
	15%	+8.0%	+13.9%	+18.4%	+20.3%	+22.5%	–	–	–	–
	20%	+7.6%	+12.1%	+16.1%	+18.7%	+21.0%	+23.1%	+24.6%	–	–
	25%	+7.9%	+13.2%	+16.5%	+19.4%	+21.7%	+23.5%	+25.5%	+27.1%	–
	100%	+8.0%	+13.2%	+16.8%	+19.9%	+22.2%	+24.1%	+26.1%	+27.3%	+28.3%
LDA-cBCI	5%	+5.5%	+11.2%	+15.1%	–	–	–	–	–	–
	10%	+9.1%	+13.1%	+17.1%	+19.4%	–	–	–	–	–
	15%	+9.1%	+15.1%	+19.0%	+21.6%	+23.6%	–	–	–	–
	20%	+9.0%	+14.4%	+18.4%	+21.1%	+23.4%	+25.4%	+26.9%	–	–
	25%	+9.1%	+15.1%	+19.0%	+21.9%	+24.1%	+26.2%	+27.6%	+29.2%	–
	100%	+9.5%	+15.7%	+20.1%	+23.0%	+25.5%	+27.4%	+29.0%	+30.5%	+31.3%

Median improvements over the average participant in the group when using collaborative BCIs for target detection at difficulty level 7, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	-6.1%	-5.9%	-2.4%	-0.4%	–	–	–	–	–
	10%	-5.9%	-4.1%	-2.6%	-0.7%	-1.8%	+0.1%	+3.0%	+2.1%	–
	15%	-4.9%	-1.9%	-0.1%	+1.4%	+2.7%	+3.9%	+5.3%	+7.6%	+8.5%
	20%	-4.9%	-1.9%	-0.1%	+1.4%	+2.7%	+3.9%	+5.3%	+7.6%	+8.5%
	25%	-4.9%	-1.9%	-0.1%	+1.4%	+2.7%	+3.9%	+5.3%	+7.6%	+8.5%
	100%	-4.9%	-1.9%	-0.1%	+1.4%	+2.7%	+3.9%	+5.3%	+7.6%	+8.5%
MC-cBCI	5%	-5.4%	-0.5%	+2.4%	+4.3%	–	–	–	–	–
	10%	-4.5%	-2.2%	-0.0%	+1.1%	+2.5%	+3.6%	+5.0%	+6.5%	–
	15%	-4.2%	-0.6%	+2.2%	+4.8%	+7.3%	+9.0%	+10.2%	+11.0%	+12.0%
	20%	-4.2%	-0.6%	+2.2%	+4.8%	+7.3%	+9.0%	+10.2%	+11.0%	+12.0%
	25%	-4.2%	-0.6%	+2.2%	+4.8%	+7.3%	+9.0%	+10.2%	+11.0%	+12.0%
	100%	-4.2%	-0.6%	+2.2%	+4.8%	+7.3%	+9.0%	+10.2%	+11.0%	+12.0%
LDA-cBCI	5%	-4.3%	+0.1%	+3.5%	+4.8%	–	–	–	–	–
	10%	-4.0%	-1.7%	-0.2%	+1.3%	+2.4%	+3.5%	+4.3%	+5.7%	–
	15%	-3.5%	-0.0%	+2.9%	+6.0%	+8.9%	+10.1%	+10.8%	+11.6%	+12.4%
	20%	-3.5%	-0.0%	+2.9%	+6.0%	+8.9%	+10.1%	+10.8%	+11.6%	+12.4%
	25%	-3.5%	-0.0%	+2.9%	+6.0%	+8.9%	+10.1%	+10.8%	+11.6%	+12.4%
	100%	-3.5%	-0.0%	+2.9%	+6.0%	+8.9%	+10.1%	+10.8%	+11.6%	+12.4%

C.1.3 Comparison with $\max(AUC_1, AUC_2, \dots, AUC_r)$

This section contains, for the target detection system, the median improvements (in percentage) when using collaborative BCIs over the AUC performance of the best participant of the group, for each level and group size.

Median improvements over the best participant in the group when using collaborative BCIs for target detection at difficulty level 1, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	-1.3%	-0.4%	-0.0%	-0.2%	—	—	—	—	—	—
	10%	-0.9%	-0.2%	-0.4%	-0.4%	-0.4%	-0.0%	—	—	—	—
	15%	-1.4%	-0.6%	-0.5%	-0.4%	-0.4%	-0.0%	—	—	—	—
	20%	-2.0%	-1.8%	-0.9%	-0.5%	-0.3%	-0.2%	+0.2%	+0.2%	—	—
	25%	-2.8%	-2.3%	-1.6%	-0.8%	-0.4%	-0.1%	+0.3%	+0.6%	—	—
	100%	<i>-3.1%</i>	-2.8%	-1.9%	-1.2%	-0.8%	-0.6%	-0.4%	-0.3%	-0.1%	+0.0%
MC-cBCI	5%	-0.1%	+0.9%	+1.4%	+1.5%	—	—	—	—	—	—
	10%	-0.1%	+0.9%	+1.3%	+1.2%	+1.6%	+1.6%	—	—	—	—
	15%	-0.1%	+0.9%	+1.1%	+1.2%	+1.6%	+1.6%	—	—	—	—
	20%	-1.2%	-0.3%	+0.6%	+1.1%	+1.4%	+1.6%	+1.7%	+1.9%	—	—
	25%	-1.4%	-1.2%	-0.2%	+0.5%	+1.0%	+1.4%	+1.7%	+1.9%	—	—
	100%	-1.6%	-1.4%	-0.6%	+0.1%	+0.4%	+0.7%	+0.9%	+1.0%	+1.1%	+1.2%
LDA-cBCI	5%	+0.3%	+0.9%	+1.5%	+1.6%	—	—	—	—	—	—
	10%	+0.8%	+0.9%	+1.3%	+1.3%	+1.5%	+1.7%	—	—	—	—
	15%	+0.4%	+0.9%	+1.2%	+1.3%	+1.5%	+1.7%	—	—	—	—
	20%	-0.7%	+0.3%	+0.9%	+1.2%	+1.5%	+1.6%	+1.8%	+1.9%	—	—
	25%	-0.8%	-0.1%	+0.6%	+1.0%	+1.5%	+1.7%	+1.9%	+2.3%	—	—
	100%	-0.8%	-0.3%	+0.4%	+0.7%	+1.0%	+1.2%	+1.3%	+1.4%	+1.5%	+1.5%

Median improvements over the best participant in the group when using collaborative BCIs for target detection at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	+2.0%	+2.1%	+2.7%	+3.0%	–	–	–	–	–
	10%	+0.3%	+0.9%	+1.3%	+1.4%	+2.1%	–	–	–	–
	15%	+0.6%	+0.9%	+1.3%	+1.5%	+2.5%	–	–	–	–
	20%	<i>+0.3%</i>	+0.8%	+0.9%	+1.0%	+0.9%	+1.6%	–	–	–
	25%	-0.0%	+0.3%	+0.8%	+1.2%	+1.4%	+2.2%	+2.5%	–	–
	100%	-0.9%	-1.2%	-0.6%	-0.4%	-0.1%	+0.2%	+0.2%	+0.9%	+0.7%
MC-cBCI	5%	+3.3%	+4.1%	+4.4%	+4.8%	–	–	–	–	–
	10%	+2.0%	+2.8%	+2.6%	+2.8%	+3.0%	–	–	–	–
	15%	+2.2%	+3.0%	+3.1%	+4.1%	+3.9%	–	–	–	–
	20%	+2.0%	+2.5%	+2.7%	+2.6%	+2.6%	+2.9%	–	–	–
	25%	+1.4%	+2.3%	+2.8%	+2.9%	+3.1%	+3.2%	+3.8%	–	–
	100%	+0.7%	+1.0%	+1.4%	+1.7%	+1.6%	+1.7%	+1.9%	+1.8%	+2.0%
LDA-cBCI	5%	+3.4%	+4.1%	+4.7%	+4.9%	–	–	–	–	–
	10%	+2.1%	+3.0%	+2.7%	+2.8%	+3.0%	–	–	–	–
	15%	+2.2%	+3.2%	+3.5%	+4.2%	+4.0%	–	–	–	–
	20%	+2.2%	+3.0%	+2.8%	+2.7%	+2.7%	+3.0%	–	–	–
	25%	+2.0%	+3.0%	+3.5%	+3.7%	+4.0%	+4.0%	+4.4%	–	–
	100%	+1.6%	+2.1%	+2.4%	+2.5%	+2.5%	+2.6%	+2.5%	+2.6%	+2.7%

Median improvements over the best participant in the group when using collaborative BCIs for target detection at difficulty level 3, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	+2.7%	+5.5%	+7.1%	–	–	–	–	–	–	–
	10%	-0.3%	+0.8%	+4.8%	+5.2%	–	–	–	–	–	–
	15%	-0.3%	+0.8%	+3.2%	+1.2%	+2.8%	–	–	–	–	–
	20%	-1.5%	+0.8%	+2.8%	+3.7%	+4.8%	+3.7%	–	–	–	–
	25%	<i>-1.9%</i>	-0.3%	+1.0%	+2.2%	+2.9%	+3.7%	+4.3%	+4.7%	–	–
	100%	-2.7%	-1.2%	-0.5%	+0.4%	+0.7%	+1.1%	+1.5%	+1.7%	+1.8%	+1.8%
MC-cBCI	5%	+3.8%	+8.3%	+9.9%	–	–	–	–	–	–	–
	10%	+0.5%	+2.4%	+5.8%	+7.9%	–	–	–	–	–	–
	15%	+0.1%	+2.1%	+1.9%	-0.0%	-0.2%	–	–	–	–	–
	20%	-0.5%	+1.2%	+3.5%	+5.3%	+6.2%	+6.6%	–	–	–	–
	25%	-2.0%	-0.5%	+1.1%	+2.3%	+3.6%	+4.4%	+5.4%	+5.8%	–	–
	100%	-3.1%	-1.6%	-0.9%	+0.1%	+0.7%	+1.3%	+1.6%	+1.9%	+2.6%	+2.6%
LDA-cBCI	5%	+4.7%	+8.6%	+10.3%	–	–	–	–	–	–	–
	10%	+1.2%	+3.2%	+6.3%	+8.3%	–	–	–	–	–	–
	15%	+1.3%	+3.1%	+4.4%	+3.3%	+3.3%	–	–	–	–	–
	20%	+1.1%	+3.2%	+5.3%	+7.0%	+7.7%	+7.8%	–	–	–	–
	25%	+0.8%	+2.6%	+4.3%	+5.6%	+6.6%	+7.1%	+7.3%	+7.8%	–	–
	100%	+0.7%	+2.0%	+3.6%	+4.5%	+5.0%	+5.3%	+5.5%	+5.6%	+5.8%	+5.8%

Median improvements over the best participant in the group when using collaborative BCIs for target detection at difficulty level 4, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	+0.2%	+3.9%	+5.7%	+7.8%	—	—	—	—	—
	10%	+0.1%	+3.7%	+6.0%	+7.8%	+8.5%	—	—	—	—
	15%	<i>-0.6%</i>	+2.4%	+5.4%	+7.0%	+8.2%	—	—	—	—
	20%	-0.6%	+2.4%	+5.3%	+6.7%	+8.0%	+9.2%	—	—	—
	25%	-1.4%	+0.1%	+2.7%	+3.2%	+6.1%	+7.7%	—	—	—
	100%	-3.5%	<i>-3.0%</i>	-0.7%	+0.8%	+1.7%	+2.7%	+4.1%	+4.5%	+6.0%
MC-cBCI	5%	+3.1%	+6.3%	+8.6%	+10.1%	—	—	—	—	—
	10%	+2.8%	+6.4%	+8.6%	+10.0%	+10.9%	—	—	—	—
	15%	+2.0%	+5.6%	+7.7%	+9.0%	+10.4%	—	—	—	—
	20%	+1.0%	+5.1%	+6.9%	+8.3%	+9.4%	+10.2%	—	—	—
	25%	-2.6%	+1.0%	+4.1%	+6.2%	+7.8%	+8.9%	—	—	—
	100%	-4.3%	-2.7%	-0.4%	+1.5%	+3.0%	+4.4%	+5.4%	+5.5%	+6.6%
LDA-cBCI	5%	+3.6%	+6.9%	+9.3%	+10.9%	—	—	—	—	—
	10%	+3.5%	+6.9%	+9.3%	+10.5%	+11.4%	—	—	—	—
	15%	+2.4%	+6.3%	+8.4%	+10.4%	+11.3%	—	—	—	—
	20%	+2.0%	+5.9%	+7.9%	+9.4%	+10.4%	+11.2%	—	—	—
	25%	-0.6%	+3.0%	+6.4%	+8.1%	+9.9%	+10.8%	—	—	—
	100%	-2.1%	+2.1%	+3.7%	+5.5%	+6.8%	+7.7%	+8.7%	+8.9%	+9.6%

Median improvements over the best participant in the group when using collaborative BCIs for target detection at difficulty level 5, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size							
		2	3	4	5	6	7	8	9
SC-cBCI	5%	-7.0%	-8.7%	-11.7%	–	–	–	–	–
	10%	-6.3%	-5.5%	-3.7%	-4.7%	-4.0%	-0.1%	–	–
	15%	-5.8%	-5.8%	-6.0%	-7.0%	-6.7%	-6.6%	-6.0%	–
	20%	-6.3%	-6.1%	-6.9%	-6.9%	-6.9%	-6.7%	-6.9%	-6.3%
	25%	-6.3%	-6.1%	-6.9%	-6.9%	-6.9%	-6.7%	-6.9%	-6.3%
	100%	-6.3%	-6.1%	-6.9%	-6.9%	-6.9%	-6.7%	-6.9%	-6.3%
MC-cBCI	5%	-2.7%	-7.8%	-10.1%	–	–	–	–	–
	10%	-5.3%	-5.9%	-5.7%	-5.5%	-5.6%	-4.8%	–	–
	15%	-5.4%	-6.7%	-7.0%	-7.7%	-7.8%	-7.1%	-7.0%	–
	20%	-6.3%	-7.7%	-8.4%	-8.2%	-8.1%	-7.9%	-7.8%	-7.0%
	25%	-6.3%	-7.7%	-8.4%	-8.2%	-8.1%	-7.9%	-7.8%	-7.0%
	100%	-6.3%	-7.7%	-8.4%	-8.2%	-8.1%	-7.9%	-7.8%	-7.0%
LDA-cBCI	5%	-3.3%	-7.0%	-9.6%	–	–	–	–	–
	10%	-3.4%	-3.6%	-2.4%	-1.7%	-0.3%	+0.4%	–	–
	15%	-3.7%	-3.9%	-3.9%	-3.9%	-4.0%	-3.8%	-2.6%	–
	20%	-4.1%	-4.2%	-3.9%	-3.3%	-2.8%	-2.1%	-1.7%	<i>-1.4%</i>
	25%	-4.1%	-4.2%	-3.9%	-3.3%	-2.8%	-2.1%	-1.7%	<i>-1.4%</i>
	100%	-4.1%	-4.2%	-3.9%	-3.3%	-2.8%	-2.1%	-1.7%	<i>-1.4%</i>

Median improvements over the best participant in the group when using collaborative BCIs for target detection at difficulty level 6, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	-1.3%	+2.2%	+5.8%	—	—	—	—	—	—
	10%	+0.4%	+3.0%	+5.3%	+4.9%	—	—	—	—	—
	15%	+1.4%	<i>+2.6%</i>	+5.0%	+5.3%	+6.2%	—	—	—	—
	20%	-1.1%	+0.0%	+0.8%	+1.7%	+1.6%	+0.8%	+2.6%	—	—
	25%	-1.1%	+0.5%	+1.2%	+2.6%	+1.9%	+1.4%	+2.4%	+1.9%	—
	100%	-1.2%	-0.0%	+0.5%	+1.1%	+1.5%	+1.6%	+3.0%	+3.3%	+4.1%
MC-cBCI	5%	-1.5%	+2.5%	+4.8%	—	—	—	—	—	—
	10%	+3.2%	+7.1%	+6.3%	+6.5%	—	—	—	—	—
	15%	+2.1%	+3.7%	+4.8%	+6.5%	+7.6%	—	—	—	—
	20%	+0.6%	+1.3%	+1.8%	+1.9%	+2.3%	+2.6%	+3.0%	—	—
	25%	+0.6%	+1.7%	+2.7%	+3.0%	+3.1%	+2.8%	<i>+2.6%</i>	+4.1%	—
	100%	+0.1%	+0.1%	+0.5%	+1.5%	+1.9%	+2.6%	+3.2%	+3.8%	+4.6%
LDA-cBCI	5%	-1.0%	+3.0%	+5.2%	—	—	—	—	—	—
	10%	+4.0%	<i>+7.2%</i>	<i>+6.7%</i>	+7.0%	—	—	—	—	—
	15%	+3.4%	<i>+4.8%</i>	+6.1%	+7.4%	+8.6%	—	—	—	—
	20%	+2.0%	+3.0%	+3.6%	+4.2%	+4.2%	+4.1%	+4.9%	—	—
	25%	+2.0%	+3.7%	<i>+4.5%</i>	+4.8%	+5.0%	+4.7%	+5.0%	+5.8%	—
	100%	+1.2%	+2.6%	+3.4%	+4.2%	+4.7%	+5.3%	+5.9%	+6.1%	+7.1%

Median improvements over the best participant in the group when using collaborative BCIs for target detection at difficulty level 7, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	-7.7%	-7.7%	-4.6%	-2.1%	–	–	–	–	–
	10%	-9.0%	-9.4%	-8.7%	-7.7%	-8.6%	-6.8%	-5.2%	-5.7%	–
	15%	-8.5%	-8.8%	-6.5%	-5.9%	-5.2%	-3.9%	-3.0%	-1.0%	-0.6%
	20%	-8.5%	-8.8%	-6.5%	-5.9%	-5.2%	-3.9%	-3.0%	-1.0%	-0.6%
	25%	-8.5%	-8.8%	-6.5%	-5.9%	-5.2%	-3.9%	-3.0%	-1.0%	-0.6%
	100%	-8.5%	-8.8%	-6.5%	-5.9%	-5.2%	-3.9%	-3.0%	-1.0%	-0.6%
MC-cBCI	5%	-5.6%	-2.0%	+1.1%	+2.5%	–	–	–	–	–
	10%	-9.1%	-7.4%	-6.8%	-5.4%	-5.0%	-3.4%	-3.5%	-1.6%	–
	15%	-8.8%	-5.5%	-4.1%	-2.7%	-1.5%	-0.1%	+1.3%	+1.9%	+2.7%
	20%	-8.8%	-5.5%	-4.1%	-2.7%	-1.5%	-0.1%	+1.3%	+1.9%	+2.7%
	25%	-8.8%	-5.5%	-4.1%	-2.7%	-1.5%	-0.1%	+1.3%	+1.9%	+2.7%
	100%	-8.8%	-5.5%	-4.1%	-2.7%	-1.5%	-0.1%	+1.3%	+1.9%	+2.7%
LDA-cBCI	5%	-4.8%	-1.1%	+2.0%	+3.0%	–	–	–	–	–
	10%	-8.6%	-7.1%	-6.6%	-5.3%	-5.0%	-3.7%	-3.2%	-2.3%	–
	15%	-7.7%	-4.9%	-3.0%	-2.1%	-0.3%	+0.9%	+1.9%	+2.4%	+3.0%
	20%	-7.7%	-4.9%	-3.0%	-2.1%	-0.3%	+0.9%	+1.9%	+2.4%	+3.0%
	25%	-7.7%	-4.9%	-3.0%	-2.1%	-0.3%	+0.9%	+1.9%	+2.4%	+3.0%
	100%	-7.7%	-4.9%	-3.0%	-2.1%	-0.3%	+0.9%	+1.9%	+2.4%	+3.0%

C.2 Group Member Selection on the Target Localisation System

C.2.1 Median AUCs with Group Member Selection

This section includes the tables for the target localisation system, which report the median AUC performance for each level and group size, as a function of the dissimilarity threshold, as well as the percentages of groups that are accepted for each level and value of δ .

Median AUC values for the three types of cBCIs for left vs right classification of targets for difficulty level 1, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	0.66	0.76	–	–	–	–	–	–	–	–
	10%	0.68	0.75	0.80	0.73	–	–	–	–	–	–
	15%	0.68	0.74	0.76	0.77	0.78	0.81	–	–	–	–
	20%	0.68	0.74	0.76	0.79	0.81	0.84	0.85	0.86	–	–
	25%	0.68	0.72	0.75	0.78	0.80	0.82	0.83	0.86	–	–
	100%	0.68	0.72	0.75	0.77	0.79	0.81	0.82	0.81	0.83	0.85
MC-cBCI	5%	0.84	0.90	–	–	–	–	–	–	–	–
	10%	0.84	0.89	0.92	0.94	–	–	–	–	–	–
	15%	0.83	0.88	0.91	0.93	0.94	0.96	–	–	–	–
	20%	0.84	0.88	0.91	0.93	0.94	0.95	0.95	0.96	–	–
	25%	0.84	0.89	0.91	0.93	0.94	0.96	0.97	0.98	–	–
	100%	0.84	0.88	0.90	0.92	0.94	0.95	0.96	0.97	0.97	0.97
LDA-cBCI	5%	0.85	0.90	–	–	–	–	–	–	–	–
	10%	0.85	0.90	0.91	0.93	–	–	–	–	–	–
	15%	0.85	0.89	0.91	0.94	0.95	0.96	–	–	–	–
	20%	0.85	0.89	0.91	0.93	0.95	0.96	0.96	0.96	–	–
	25%	0.85	0.89	0.91	0.93	0.95	0.96	0.96	0.96	–	–
	100%	0.85	0.89	0.91	0.92	0.94	0.95	0.96	0.96	0.97	0.97

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 1.

δ	Group size									
	2	3	4	5	6	7	8	9	10	11
5%	25%	3%	0%	0%	0%	0%	0%	0%	0%	0%
10%	47%	15%	3%	0%	0%	0%	0%	0%	0%	0%
15%	70%	43%	23%	11%	4%	0%	0%	0%	0%	0%
20%	85%	66%	47%	31%	19%	11%	5%	1%	0%	0%
25%	92%	80%	65%	50%	34%	21%	10%	3%	0%	0%

Median AUC values for the three types of cBCIs for left vs right classification of targets for difficulty level 3, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	0.61	0.67	0.75	0.86	–	–	–	–	–	–
	10%	0.65	0.68	0.71	0.74	0.77	0.79	0.79	–	–	–
	15%	0.63	0.66	0.69	0.73	0.75	0.78	0.79	–	–	–
	20%	0.63	0.63	0.66	0.69	0.70	0.70	0.73	0.75	–	–
	25%	0.64	0.65	0.66	0.67	0.69	0.70	0.73	0.73	0.75	0.75
	100%	0.64	0.65	0.66	0.67	0.69	0.70	0.73	0.73	0.75	0.75
MC-cBCI	5%	0.70	0.70	0.69	0.72	–	–	–	–	–	–
	10%	0.69	0.73	0.72	0.73	0.74	0.77	0.79	–	–	–
	15%	0.69	0.73	0.75	0.76	0.76	0.77	0.79	–	–	–
	20%	0.70	0.74	0.76	0.78	0.78	0.80	0.82	0.84	–	–
	25%	0.73	0.75	0.78	0.80	0.82	0.83	0.85	0.86	0.87	0.88
	100%	0.73	0.75	0.78	0.80	0.82	0.83	0.85	0.86	0.87	0.88
LDA-cBCI	5%	0.73	0.71	0.70	0.72	–	–	–	–	–	–
	10%	0.72	0.73	0.73	0.74	0.75	0.77	0.78	–	–	–
	15%	0.71	0.74	0.74	0.74	0.75	0.77	0.78	–	–	–
	20%	0.72	0.75	0.76	0.76	0.77	0.78	0.80	0.83	–	–
	25%	0.73	0.75	0.78	0.79	0.79	0.80	0.81	0.83	0.84	0.85
	100%	0.73	0.75	0.78	0.79	0.79	0.80	0.81	0.83	0.84	0.85

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 3.

δ	Group size									
	2	3	4	5	6	7	8	9	10	11
5%	30%	7%	1%	0%	0%	0%	0%	0%	0%	0%
10%	58%	34%	21%	12%	6%	2%	0%	0%	0%	0%
15%	80%	55%	33%	17%	7%	2%	0%	0%	0%	0%
20%	90%	76%	59%	42%	27%	15%	6%	1%	0%	0%
25%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Median AUC values for the three types of cBCIs for left vs right classification of targets for difficulty level 4, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	0.62	0.59	—	—	—	—	—	—	—
	10%	0.71	0.82	0.85	—	—	—	—	—	—
	15%	0.71	0.81	0.87	0.90	—	—	—	—	—
	20%	0.69	0.79	0.83	0.86	0.90	0.87	—	—	—
	25%	0.68	0.72	0.81	0.84	0.90	0.87	—	—	—
	100%	0.64	0.70	0.72	0.75	0.77	0.77	0.78	0.81	0.83
MC-cBCI	5%	0.64	0.61	—	—	—	—	—	—	—
	10%	0.78	0.83	0.87	—	—	—	—	—	—
	15%	0.77	0.81	0.85	0.89	—	—	—	—	—
	20%	0.73	0.78	0.80	0.82	0.84	0.85	—	—	—
	25%	0.72	0.73	0.77	0.81	0.84	0.85	—	—	—
	100%	0.69	0.72	0.76	0.77	0.78	0.80	0.79	0.80	0.81
LDA-cBCI	5%	0.66	0.63	—	—	—	—	—	—	—
	10%	0.77	0.83	0.87	—	—	—	—	—	—
	15%	0.76	0.81	0.85	0.89	—	—	—	—	—
	20%	0.73	0.78	0.81	0.83	0.84	0.85	—	—	—
	25%	0.72	0.76	0.79	0.82	0.84	0.85	—	—	—
	100%	0.72	0.76	0.77	0.79	0.80	0.81	0.80	0.81	0.80

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 4.

δ	Group size								
	2	3	4	5	6	7	8	9	10
5%	20%	1%	0%	0%	0%	0%	0%	0%	0%
10%	35%	8%	0%	0%	0%	0%	0%	0%	0%
15%	46%	19%	6%	1%	0%	0%	0%	0%	0%
20%	62%	33%	17%	8%	3%	0%	0%	0%	0%
25%	73%	45%	23%	10%	3%	0%	0%	0%	0%

Median AUC values for the three types of cBCIs for left vs right classification of targets for difficulty level 5, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size							
		2	3	4	5	6	7	8	9
SC-cBCI	5%	0.57	0.56	–	–	–	–	–	–
	10%	0.55	0.56	0.46	0.41	–	–	–	–
	15%	0.55	0.53	0.52	0.47	0.44	0.40	–	–
	20%	0.54	0.50	0.52	0.50	0.51	0.49	0.53	–
	25%	0.54	0.51	0.51	0.49	0.48	0.48	0.49	0.52
	100%	0.54	0.51	0.51	0.49	0.48	0.48	0.49	0.52
MC-cBCI	5%	0.74	0.78	–	–	–	–	–	–
	10%	0.67	0.70	0.68	0.66	–	–	–	–
	15%	0.67	0.70	0.68	0.67	0.69	0.67	–	–
	20%	0.64	0.68	0.71	0.73	0.76	0.79	0.81	–
	25%	0.64	0.68	0.71	0.73	0.75	0.77	0.78	0.81
	100%	0.64	0.68	0.71	0.73	0.75	0.77	0.78	0.81
LDA-cBCI	5%	0.74	0.77	–	–	–	–	–	–
	10%	0.69	0.73	0.71	0.69	–	–	–	–
	15%	0.69	0.70	0.72	0.73	0.72	0.71	–	–
	20%	0.69	0.73	0.76	0.79	0.82	0.83	0.84	–
	25%	0.69	0.73	0.76	0.79	0.82	0.83	0.83	0.87
	100%	0.69	0.73	0.76	0.79	0.82	0.83	0.83	0.87

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 5.

δ	Group size							
	2	3	4	5	6	7	8	9
5%	27%	3%	0%	0%	0%	0%	0%	0%
10%	63%	30%	11%	2%	0%	0%	0%	0%
15%	83%	60%	38%	21%	9%	2%	0%	0%
20%	97%	91%	83%	72%	58%	41%	22%	0%
25%	100%	100%	100%	100%	100%	100%	100%	100%

Median AUC values for the three types of cBCIs for left vs right classification of targets for difficulty level 6, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	0.60	0.62	0.64	0.65	0.68	–	–	–	–
	10%	0.61	0.63	0.62	0.61	0.60	0.67	–	–	–
	15%	0.61	0.64	0.64	0.67	0.66	0.69	0.64	–	–
	20%	0.61	0.64	0.65	0.67	0.67	0.68	0.68	0.71	–
	25%	0.61	0.64	0.66	0.68	0.68	0.68	0.68	0.67	0.66
	100%	0.61	0.64	0.66	0.68	0.68	0.68	0.68	0.67	0.66
MC-cBCI	5%	0.77	0.79	0.83	0.84	0.88	–	–	–	–
	10%	0.76	0.79	0.81	0.83	0.84	0.86	–	–	–
	15%	0.77	0.80	0.83	0.85	0.87	0.88	0.90	–	–
	20%	0.77	0.80	0.83	0.85	0.87	0.88	0.90	0.93	–
	25%	0.79	0.80	0.83	0.85	0.87	0.89	0.90	0.91	0.93
	100%	0.79	0.80	0.83	0.85	0.87	0.89	0.90	0.91	0.93
LDA-cBCI	5%	0.79	0.84	0.87	0.89	0.91	–	–	–	–
	10%	0.80	0.83	0.85	0.88	0.90	0.91	–	–	–
	15%	0.81	0.85	0.89	0.91	0.92	0.95	0.96	–	–
	20%	0.81	0.85	0.89	0.91	0.92	0.94	0.97	0.97	–
	25%	0.82	0.86	0.89	0.91	0.93	0.94	0.96	0.97	0.97
	100%	0.82	0.86	0.89	0.91	0.93	0.94	0.96	0.97	0.97

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 6.

δ	Group size								
	2	3	4	5	6	7	8	9	10
5%	40%	16%	7%	2%	0%	0%	0%	0%	0%
10%	66%	42%	26%	14%	6%	1%	0%	0%	0%
15%	91%	76%	59%	42%	26%	13%	4%	0%	0%
20%	95%	87%	76%	63%	50%	35%	22%	10%	0%
25%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Median AUC values for the three types of cBCIs for left vs right classification of targets for difficulty level 7, as a function of group size and the dissimilarity-index threshold δ .

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	0.48	0.44	0.41	0.39	0.37	–	–	–	–
	10%	0.50	0.47	0.47	0.43	0.42	0.47	0.34	–	–
	15%	0.54	0.51	0.48	0.43	0.42	0.47	0.34	–	–
	20%	0.54	0.53	0.51	0.52	0.52	0.53	0.52	0.42	0.34
	25%	0.54	0.53	0.51	0.52	0.52	0.53	0.52	0.42	0.34
	100%	0.54	0.53	0.51	0.52	0.52	0.53	0.52	0.42	0.34
MC-cBCI	5%	0.57	0.54	0.53	0.55	0.55	–	–	–	–
	10%	0.56	0.54	0.52	0.54	0.53	0.54	0.53	–	–
	15%	0.58	0.56	0.57	0.55	0.54	0.54	0.53	–	–
	20%	0.58	0.63	0.64	0.67	0.68	0.70	0.71	0.70	0.70
	25%	0.58	0.63	0.64	0.67	0.68	0.70	0.71	0.70	0.70
	100%	0.58	0.63	0.64	0.67	0.68	0.70	0.71	0.70	0.70
LDA-cBCI	5%	0.54	0.52	0.49	0.52	0.53	–	–	–	–
	10%	0.56	0.53	0.53	0.54	0.53	0.49	0.48	–	–
	15%	0.57	0.56	0.56	0.55	0.53	0.49	0.48	–	–
	20%	0.59	0.62	0.63	0.66	0.66	0.67	0.67	0.67	0.70
	25%	0.59	0.62	0.63	0.66	0.66	0.67	0.67	0.67	0.70
	100%	0.59	0.62	0.63	0.66	0.66	0.67	0.67	0.67	0.70

Percentages of groups that are accepted by the selection mechanism for the target vs non-target discrimination task for different group sizes and values of the dissimilarity-index threshold δ at difficulty level 7.

δ	Group size								
	2	3	4	5	6	7	8	9	10
5%	44%	19%	7%	2%	0%	0%	0%	0%	0%
10%	71%	48%	33%	22%	13%	6%	2%	0%	0%
15%	82%	60%	40%	24%	13%	6%	2%	0%	0%
20%	100%	100%	100%	100%	100%	100%	100%	100%	100%
25%	100%	100%	100%	100%	100%	100%	100%	100%	100%

C.2.2 Comparison with $\text{avg}(AUC_1, AUC_2, \dots, AUC_r)$

The tables included in this section report, for the target localisation system, the median improvements (in percentage) when using collaborative BCIs over the average AUC performance of the group, for each level and group size.

Median improvements over the average participant in the group when using collaborative BCIs for target localisation at difficulty level 1, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	-9.1%	-4.3%	–	–	–	–	–	–	–	–
	10%	-9.9%	-3.7%	-1.7%	-1.7%	–	–	–	–	–	–
	15%	-8.4%	-2.0%	+2.0%	+5.2%	+4.1%	+1.0%	–	–	–	–
	20%	-9.8%	-4.3%	+0.2%	+1.8%	+4.7%	+6.1%	+8.0%	+6.6%	–	–
	25%	-10.6%	-4.7%	-1.5%	+1.0%	+3.1%	+4.1%	+4.6%	+4.2%	–	–
	100%	-10.2%	-4.5%	-1.0%	+1.3%	+3.4%	+4.2%	+4.9%	+5.0%	+8.2%	+6.4%
MC-cBCI	5%	+4.1%	+13.5%	–	–	–	–	–	–	–	–
	10%	+3.7%	+7.7%	+9.9%	+13.5%	–	–	–	–	–	–
	15%	+4.6%	+10.5%	+14.9%	+18.0%	+19.2%	+21.5%	–	–	–	–
	20%	+3.7%	+9.2%	+14.5%	+17.9%	+20.0%	+21.7%	+22.8%	+26.3%	–	–
	25%	+2.8%	+7.9%	+12.3%	+15.8%	+18.0%	+20.1%	+22.3%	+25.3%	–	–
	100%	+2.4%	+7.9%	+11.8%	+15.3%	+17.9%	+20.2%	+21.9%	+23.2%	+24.3%	+25.6%
LDA-cBCI	5%	+2.7%	+5.4%	–	–	–	–	–	–	–	–
	10%	+2.8%	+9.0%	+10.6%	+12.3%	–	–	–	–	–	–
	15%	+4.7%	+10.7%	+15.4%	+18.9%	+20.0%	+22.2%	–	–	–	–
	20%	+3.4%	+9.6%	+14.7%	+17.8%	+20.4%	+22.1%	+23.2%	+24.3%	–	–
	25%	+2.1%	+9.1%	+13.1%	+16.4%	+19.2%	+20.9%	+21.4%	+23.6%	–	–
	100%	+3.4%	+9.3%	+13.2%	+16.4%	+19.2%	+21.4%	+23.1%	+24.5%	+25.2%	+25.3%

Median improvements over the average participant in the group when using collaborative BCIs for target localisation at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	-8.9%	-8.5%	—	—	—	—	—	—	—
	10%	-5.2%	-3.3%	-1.7%	+4.1%	—	—	—	—	—
	15%	-5.2%	-2.7%	-0.1%	+0.1%	-2.4%	—	—	—	—
	20%	-3.5%	<i>-1.4%</i>	+0.8%	+3.5%	+5.7%	+7.6%	—	—	—
	25%	-2.8%	-0.3%	+2.5%	+5.9%	+6.9%	+9.2%	+10.3%	+14.1%	—
	100%	-2.5%	+0.0%	+3.3%	+6.4%	+8.2%	+10.1%	+11.5%	+10.4%	+13.9%
MC-cBCI	5%	+1.4%	+6.1%	—	—	—	—	—	—	—
	10%	+2.7%	+5.5%	+9.8%	+12.8%	—	—	—	—	—
	15%	+2.7%	+7.3%	+10.7%	+13.3%	+17.1%	—	—	—	—
	20%	+3.0%	+7.6%	+10.4%	+12.4%	+13.7%	+15.2%	—	—	—
	25%	+3.5%	+8.0%	+11.4%	+14.7%	+17.4%	+20.5%	+22.2%	+22.9%	—
	100%	+3.9%	+8.9%	+12.3%	+15.6%	+17.7%	+20.1%	+21.5%	+22.8%	+23.9%
LDA-cBCI	5%	+0.2%	+4.1%	—	—	—	—	—	—	—
	10%	+3.3%	+5.8%	+9.2%	+12.5%	—	—	—	—	—
	15%	+4.0%	+7.6%	+10.2%	+12.4%	+18.3%	—	—	—	—
	20%	+4.0%	+8.1%	+10.5%	+12.3%	+13.9%	+16.1%	—	—	—
	25%	+4.4%	+9.9%	+13.8%	+16.6%	+19.5%	+21.9%	+23.7%	+24.9%	—
	100%	+4.7%	+11.2%	+14.8%	+17.1%	+19.1%	+20.1%	+21.5%	+22.2%	+23.4%

Median improvements over the average participant in the group when using collaborative BCIs for target localisation at difficulty level 3, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	+7.0%	+14.5%	+11.6%	+13.1%	—	—	—	—	—	—
	10%	+10.3%	+14.9%	+17.1%	+20.6%	+21.7%	+23.1%	+28.1%	—	—	—
	15%	+8.4%	+10.9%	+13.9%	+17.5%	+19.7%	+21.8%	+28.1%	—	—	—
	20%	+5.4%	+9.3%	+12.6%	+14.9%	+15.7%	+19.4%	+19.4%	+16.3%	—	—
	25%	+3.0%	+7.7%	+10.1%	+12.0%	+12.0%	+13.8%	+15.6%	+14.5%	+16.6%	+14.7%
	100%	+3.0%	+7.7%	+10.1%	+12.0%	+12.0%	+13.8%	+15.6%	+14.5%	+16.6%	+14.7%
MC-cBCI	5%	+21.2%	+33.3%	+42.1%	+45.8%	—	—	—	—	—	—
	10%	+20.8%	+28.8%	+34.8%	+38.2%	+40.4%	+43.2%	+44.5%	—	—	—
	15%	+18.2%	+25.4%	+32.8%	+35.9%	+39.6%	+42.5%	+44.5%	—	—	—
	20%	+16.1%	+24.0%	+31.1%	+35.2%	+39.6%	+42.5%	+45.7%	+46.6%	—	—
	25%	+15.5%	+23.0%	+27.1%	+31.3%	+34.4%	+37.4%	+39.8%	+41.5%	+42.5%	+45.5%
	100%	+15.5%	+23.0%	+27.1%	+31.3%	+34.4%	+37.4%	+39.8%	+41.5%	+42.5%	+45.5%
LDA-cBCI	5%	+22.9%	+33.6%	+41.7%	+47.2%	—	—	—	—	—	—
	10%	+20.1%	+29.6%	+34.9%	+38.6%	+41.3%	+43.1%	+45.3%	—	—	—
	15%	+18.8%	+28.0%	+33.7%	+37.6%	+41.3%	+43.3%	+45.3%	—	—	—
	20%	+18.0%	+26.0%	+31.4%	+35.8%	+40.2%	+42.6%	+45.3%	+46.5%	—	—
	25%	+17.2%	+23.9%	+28.6%	+33.0%	+36.2%	+38.8%	+41.3%	+43.1%	+43.2%	+44.7%
	100%	+17.2%	+23.9%	+28.6%	+33.0%	+36.2%	+38.8%	+41.3%	+43.1%	+43.2%	+44.7%

Median improvements over the average participant in the group when using collaborative BCIs for target localisation at difficulty level 4, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	-7.5%	-6.7%	—	—	—	—	—	—	—
	10%	-4.8%	<i>+2.7%</i>	<i>+7.1%</i>	—	—	—	—	—	—
	15%	-4.8%	+4.7%	+9.9%	<i>+14.2%</i>	—	—	—	—	—
	20%	-5.4%	<i>+2.3%</i>	+7.3%	+14.0%	+13.0%	<i>+22.6%</i>	—	—	—
	25%	-5.2%	<i>+0.3%</i>	+4.9%	+9.7%	+12.5%	<i>+22.6%</i>	—	—	—
	100%	-5.3%	-2.8%	+1.9%	+5.6%	+8.7%	+10.9%	+10.8%	+16.5%	<i>+20.0%</i>
MC-cBCI	5%	<i>+4.5%</i>	<i>+5.7%</i>	—	—	—	—	—	—	—
	10%	<i>+9.0%</i>	+15.7%	<i>+18.3%</i>	—	—	—	—	—	—
	15%	<i>+10.7%</i>	+16.2%	+18.4%	<i>+22.1%</i>	—	—	—	—	—
	20%	<i>+7.1%</i>	<i>+15.4%</i>	+18.9%	+22.0%	+23.2%	<i>+26.0%</i>	—	—	—
	25%	<i>+7.3%</i>	<i>+14.5%</i>	+17.3%	+21.2%	+23.1%	<i>+26.0%</i>	—	—	—
	100%	<i>+7.6%</i>	+14.7%	+18.0%	+21.1%	+23.4%	+25.8%	+27.0%	+29.5%	<i>+31.8%</i>
LDA-cBCI	5%	<i>+5.2%</i>	<i>+3.1%</i>	—	—	—	—	—	—	—
	10%	<i>+9.3%</i>	+15.6%	<i>+18.3%</i>	—	—	—	—	—	—
	15%	<i>+10.1%</i>	+16.6%	+18.6%	<i>+23.0%</i>	—	—	—	—	—
	20%	<i>+8.7%</i>	+17.0%	+19.8%	+23.4%	+24.5%	<i>+27.5%</i>	—	—	—
	25%	<i>+8.9%</i>	+17.6%	+20.8%	+23.0%	+24.3%	<i>+27.5%</i>	—	—	—
	100%	<i>+10.7%</i>	+18.9%	+22.6%	+25.4%	+27.5%	+29.4%	+31.2%	+33.2%	<i>+34.7%</i>

Median improvements over the average participant in the group when using collaborative BCIs for target localisation at difficulty level 5, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size							
		2	3	4	5	6	7	8	9
SC-cBCI	5%	-11.2%	-14.7%	—	—	—	—	—	—
	10%	-13.5%	-16.5%	-17.3%	-16.6%	—	—	—	—
	15%	-13.5%	-12.4%	-12.2%	-11.9%	-14.2%	-12.1%	—	—
	20%	-14.9%	-14.7%	-12.2%	-10.8%	-10.0%	-9.1%	-10.0%	—
	25%	-14.2%	-13.6%	-11.0%	-9.9%	-7.9%	-7.6%	-7.2%	-1.5%
	100%	-14.2%	-13.6%	-11.0%	-9.9%	-7.9%	-7.6%	-7.2%	-1.5%
MC-cBCI	5%	-8.6%	-6.7%	—	—	—	—	—	—
	10%	<i>-7.3%</i>	-7.1%	-7.3%	-7.3%	—	—	—	—
	15%	-6.1%	-6.7%	-7.0%	-7.2%	-5.1%	-5.4%	—	—
	20%	-7.3%	-4.8%	-2.5%	-0.5%	+0.8%	+3.3%	+3.2%	—
	25%	-6.1%	<i>-3.9%</i>	-1.7%	+0.9%	+1.9%	+3.4%	+3.4%	+3.8%
	100%	-6.1%	<i>-3.9%</i>	-1.7%	+0.9%	+1.9%	+3.4%	+3.4%	+3.8%
LDA-cBCI	5%	-8.2%	-0.9%	—	—	—	—	—	—
	10%	-7.4%	-2.9%	-0.2%	+1.4%	—	—	—	—
	15%	<i>-4.6%</i>	-1.3%	-0.3%	<i>-0.9%</i>	-0.2%	+0.3%	—	—
	20%	-2.7%	+3.2%	+5.5%	+7.6%	+9.7%	+8.8%	+10.9%	—
	25%	-2.5%	+3.4%	+6.9%	+9.6%	+10.9%	+12.0%	+13.7%	+13.4%
	100%	-2.5%	+3.4%	+6.9%	+9.6%	+10.9%	+12.0%	+13.7%	+13.4%

Median improvements over the average participant in the group when using collaborative BCIs for target localisation at difficulty level 6, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	-13.4%	-13.2%	-8.5%	-8.6%	-14.3%	—	—	—	—
	10%	-12.0%	-12.4%	-11.2%	-10.7%	-7.0%	-6.8%	—	—	—
	15%	-17.1%	-13.3%	-13.1%	-12.1%	-8.2%	-8.0%	-4.7%	—	—
	20%	-16.9%	-13.9%	-12.8%	-12.5%	-9.9%	-8.1%	-4.9%	-2.5%	—
	25%	-16.9%	-15.3%	-14.9%	-15.8%	-14.5%	-14.7%	-15.0%	-8.6%	-5.5%
	100%	-16.9%	-15.3%	-14.9%	-15.8%	-14.5%	-14.7%	-15.0%	-8.6%	-5.5%
MC-cBCI	5%	-5.2%	-0.8%	+1.8%	+6.2%	+5.9%	—	—	—	—
	10%	-3.3%	+0.3%	+2.3%	+6.7%	+6.5%	+8.4%	—	—	—
	15%	-4.2%	+0.3%	+2.7%	+5.8%	+7.5%	+9.8%	+11.6%	—	—
	20%	-4.2%	+0.3%	+3.1%	+6.3%	+8.4%	+10.5%	+12.0%	+13.3%	—
	25%	-5.2%	-0.4%	+2.1%	+4.3%	+6.7%	+8.6%	+9.7%	+10.5%	+11.4%
	100%	-5.2%	-0.4%	+2.1%	+4.3%	+6.7%	+8.6%	+9.7%	+10.5%	+11.4%
LDA-cBCI	5%	-2.3%	+0.9%	+3.8%	+6.1%	+7.6%	—	—	—	—
	10%	-1.9%	+1.6%	+4.9%	+6.3%	+8.3%	+9.8%	—	—	—
	15%	-2.4%	+1.7%	+5.2%	+7.0%	+8.6%	+9.7%	+12.2%	—	—
	20%	-2.4%	+1.7%	+4.9%	+7.5%	+9.3%	+11.0%	+13.5%	+12.9%	—
	25%	-2.4%	+1.7%	+4.6%	+7.3%	+9.2%	+10.6%	+12.1%	+13.1%	+12.2%
	100%	-2.4%	+1.7%	+4.6%	+7.3%	+9.2%	+10.6%	+12.1%	+13.1%	+12.2%

Median improvements over the average participant in the group when using collaborative BCIs for target localisation at difficulty level 7, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs average AUC of the group). Values in italics are statistically superior at the 5% confidence level.

		Group size								
Method	δ	2	3	4	5	6	7	8	9	10
SC-cBCI	5%	-12.0%	-11.6%	-6.0%	-2.8%	-4.6%	—	—	—	—
	10%	-7.8%	-6.9%	-4.6%	-2.3%	-3.9%	+2.4%	-2.3%	—	—
	15%	-6.3%	-5.2%	-3.6%	-1.3%	-3.7%	+2.4%	-2.3%	—	—
	20%	-8.4%	-7.0%	-5.1%	-3.6%	-2.2%	-1.7%	-4.0%	-5.5%	-13.2%
	25%	-8.4%	-7.0%	-5.1%	-3.6%	-2.2%	-1.7%	-4.0%	-5.5%	-13.2%
	100%	-8.4%	-7.0%	-5.1%	-3.6%	-2.2%	-1.7%	-4.0%	-5.5%	-13.2%
MC-cBCI	5%	-3.8%	-1.2%	-0.0%	+2.3%	+7.8%	—	—	—	—
	10%	+0.9%	+6.6%	+10.6%	+12.2%	+12.1%	+14.0%	+15.7%	—	—
	15%	+1.7%	+9.2%	+11.8%	+13.4%	+12.1%	+14.0%	+15.7%	—	—
	20%	+2.9%	+9.4%	+13.5%	+16.0%	+18.0%	+19.7%	+20.9%	+22.8%	+23.6%
	25%	+2.9%	+9.4%	+13.5%	+16.0%	+18.0%	+19.7%	+20.9%	+22.8%	+23.6%
	100%	+2.9%	+9.4%	+13.5%	+16.0%	+18.0%	+19.7%	+20.9%	+22.8%	+23.6%
LDA-cBCI	5%	-4.3%	-4.9%	-0.6%	+1.2%	+4.4%	—	—	—	—
	10%	-0.8%	+4.0%	+9.3%	+10.3%	+10.7%	+9.9%	+10.1%	—	—
	15%	-0.1%	+9.4%	+11.7%	+11.1%	+11.0%	+9.9%	+10.1%	—	—
	20%	+2.8%	+9.9%	+13.7%	+15.6%	+17.7%	+20.5%	+20.9%	+20.9%	+19.2%
	25%	+2.8%	+9.9%	+13.7%	+15.6%	+17.7%	+20.5%	+20.9%	+20.9%	+19.2%
	100%	+2.8%	+9.9%	+13.7%	+15.6%	+17.7%	+20.5%	+20.9%	+20.9%	+19.2%

C.2.3 Comparison with $\max(AUC_1, AUC_2, \dots, AUC_r)$

The results in this section report, for the target localisation system, the median improvements (in percentage) when using collaborative BCIs over the AUC performance of the best participant of the group, for each level and group size.

Median improvements over the best participant in the group when using collaborative BCIs for target localisation at difficulty level 1, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	-12.2%	-7.6%	–	–	–	–	–	–	–	–
	10%	-13.2%	-9.1%	-10.2%	-10.2%	–	–	–	–	–	–
	15%	-12.3%	-11.0%	-7.3%	-5.0%	-4.9%	-7.9%	–	–	–	–
	20%	-14.0%	-11.6%	-8.3%	-6.8%	-5.5%	-4.7%	-3.2%	-4.4%	–	–
	25%	-15.0%	-11.7%	-9.9%	-8.5%	-7.6%	-7.6%	-5.8%	-6.9%	–	–
	100%	-14.3%	-11.6%	-10.0%	-8.4%	-7.4%	-7.0%	-6.0%	-6.7%	<i>-4.1%</i>	<i>-5.7%</i>
MC-cBCI	5%	+0.4%	+0.3%	–	–	–	–	–	–	–	–
	10%	-4.0%	-0.5%	+1.6%	+3.5%	–	–	–	–	–	–
	15%	-2.2%	+1.8%	+4.7%	+6.2%	+6.5%	+8.8%	–	–	–	–
	20%	-2.6%	+1.4%	+4.6%	+7.1%	+8.5%	+9.6%	+11.3%	+13.3%	–	–
	25%	-3.3%	+0.1%	+2.4%	+4.6%	+6.6%	+7.9%	+9.3%	+12.0%	–	–
	100%	-3.3%	-0.1%	+2.2%	+4.2%	+6.2%	+7.5%	+8.6%	+10.0%	+10.5%	+11.3%
LDA-cBCI	5%	+0.7%	+1.7%	–	–	–	–	–	–	–	–
	10%	-4.0%	+1.2%	+2.2%	+2.4%	–	–	–	–	–	–
	15%	-1.6%	+2.2%	+5.1%	+6.7%	+8.8%	<i>+8.5%</i>	–	–	–	–
	20%	-4.1%	+1.5%	+4.8%	+6.8%	+8.8%	+9.8%	+10.2%	+11.5%	–	–
	25%	-4.2%	+0.9%	+3.2%	+5.4%	+7.0%	+8.7%	+9.6%	+10.4%	–	–
	100%	-4.1%	+1.2%	+3.2%	+5.2%	+6.9%	+8.7%	+9.7%	+10.8%	+11.6%	+11.0%

Median improvements over the best participant in the group when using collaborative BCIs for target localisation at difficulty level 2, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	<i>-12.8%</i>	-11.4%	—	—	—	—	—	—	—
	10%	-11.0%	-10.0%	-9.0%	-4.1%	—	—	—	—	—
	15%	-12.7%	-10.9%	-10.5%	-10.0%	-13.2%	—	—	—	—
	20%	-12.6%	-10.5%	-10.1%	-7.7%	<i>-4.4%</i>	-4.0%	—	—	—
	25%	-12.5%	-11.3%	-10.5%	-9.1%	-8.0%	-6.8%	-7.1%	-5.1%	—
	100%	-12.6%	-11.5%	-10.2%	-8.2%	-7.0%	-5.9%	-5.4%	-6.4%	-4.1%
MC-cBCI	5%	<i>-1.3%</i>	+2.7%	—	—	—	—	—	—	—
	10%	-1.9%	+0.5%	+3.0%	+3.9%	—	—	—	—	—
	15%	-2.5%	-1.4%	+0.6%	+2.3%	+3.8%	—	—	—	—
	20%	-2.5%	-1.6%	-0.5%	+1.6%	+2.7%	+2.8%	—	—	—
	25%	-3.9%	-2.1%	-1.2%	+0.0%	+1.4%	+1.7%	+2.4%	+2.2%	—
	100%	-4.2%	-2.5%	-1.3%	+0.1%	+1.4%	+2.1%	+2.8%	+4.0%	+4.3%
LDA-cBCI	5%	-1.5%	+0.9%	—	—	—	—	—	—	—
	10%	-1.1%	+0.6%	+1.8%	+3.6%	—	—	—	—	—
	15%	-2.4%	-0.9%	+1.1%	+3.7%	+4.8%	—	—	—	—
	20%	-2.8%	-1.5%	+0.5%	+1.4%	+2.4%	+3.1%	—	—	—
	25%	-2.8%	-1.2%	+0.5%	+1.3%	+2.1%	+2.8%	+2.9%	+3.9%	—
	100%	-2.7%	-1.1%	+0.4%	+1.3%	+2.0%	+2.6%	+3.0%	+3.6%	+3.9%

Median improvements over the best participant in the group when using collaborative BCIs for target localisation at difficulty level 3, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size									
		2	3	4	5	6	7	8	9	10	11
SC-cBCI	5%	-5.5%	+0.7%	-5.5%	-6.1%	—	—	—	—	—	—
	10%	+3.4%	+0.6%	+1.8%	+2.7%	+3.1%	+4.2%	+7.5%	—	—	—
	15%	-0.2%	-0.2%	-1.1%	<i>+0.7%</i>	+1.3%	+3.5%	+7.5%	—	—	—
	20%	-1.7%	-1.9%	-3.0%	-2.4%	-2.5%	<i>+0.2%</i>	-1.2%	-4.2%	—	—
	25%	-5.5%	-3.2%	-4.4%	-4.5%	-5.0%	-3.5%	-3.9%	-4.5%	-3.6%	-5.2%
	100%	-5.5%	-3.2%	-4.4%	-4.5%	-5.0%	-3.5%	-3.9%	-4.5%	-3.6%	-5.2%
MC-cBCI	5%	+10.1%	+14.2%	+18.9%	+21.1%	—	—	—	—	—	—
	10%	+10.3%	+13.8%	+16.2%	+17.8%	+19.7%	+20.3%	+21.3%	—	—	—
	15%	+10.1%	+12.7%	+14.5%	+16.8%	+19.5%	+19.9%	+21.3%	—	—	—
	20%	+9.9%	+11.5%	+13.1%	+15.9%	+17.5%	+19.5%	+21.2%	+20.9%	—	—
	25%	+8.4%	+10.2%	+11.5%	+13.4%	+14.6%	+16.1%	+16.8%	+17.7%	+18.5%	+20.2%
	100%	+8.4%	+10.2%	+11.5%	+13.4%	+14.6%	+16.1%	+16.8%	+17.7%	+18.5%	+20.2%
LDA-cBCI	5%	+10.5%	+15.0%	+19.6%	+22.2%	—	—	—	—	—	—
	10%	+10.5%	+14.2%	+16.9%	+18.8%	+20.5%	+21.2%	+21.9%	—	—	—
	15%	+10.0%	+13.6%	+15.5%	+18.1%	+19.6%	+21.1%	+21.9%	—	—	—
	20%	+9.7%	+12.2%	+14.4%	+16.3%	+17.9%	+19.7%	+19.5%	+20.7%	—	—
	25%	+8.7%	+11.2%	+12.9%	+14.6%	+16.0%	+16.7%	+18.4%	+18.9%	+18.7%	+19.6%
	100%	+8.7%	+11.2%	+12.9%	+14.6%	+16.0%	+16.7%	+18.4%	+18.9%	+18.7%	+19.6%

Median improvements over the best participant in the group when using collaborative BCIs for target localisation at difficulty level 4, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	-16.4%	-12.7%	—	—	—	—	—	—	—
	10%	-14.6%	-9.6%	-8.5%	—	—	—	—	—	—
	15%	-12.8%	-7.8%	-7.0%	-6.2%	—	—	—	—	—
	20%	-12.3%	-8.3%	-5.8%	-6.5%	-5.0%	+1.2%	—	—	—
	25%	-11.8%	-11.4%	-9.2%	-6.8%	-5.5%	+1.2%	—	—	—
	100%	-12.8%	-15.1%	-13.4%	-13.1%	-11.6%	-10.6%	-13.2%	-9.0%	-7.6%
MC-cBCI	5%	-2.8%	-1.1%	—	—	—	—	—	—	—
	10%	-0.6%	<i>+0.6%</i>	+0.8%	—	—	—	—	—	—
	15%	+0.4%	+2.3%	+1.3%	-0.9%	—	—	—	—	—
	20%	-1.0%	+2.2%	+2.2%	+2.5%	+3.1%	+4.0%	—	—	—
	25%	-0.6%	+2.2%	+1.7%	+1.9%	+2.9%	+4.0%	—	—	—
	100%	-1.5%	-0.7%	-0.5%	-0.6%	-0.6%	-0.1%	-0.3%	+0.0%	+1.4%
LDA-cBCI	5%	-2.8%	-3.6%	—	—	—	—	—	—	—
	10%	-0.0%	+1.6%	+0.9%	—	—	—	—	—	—
	15%	+0.8%	+3.0%	<i>-0.1%</i>	-1.1%	—	—	—	—	—
	20%	-0.4%	+3.3%	+3.7%	+3.3%	+4.2%	+5.2%	—	—	—
	25%	+0.1%	+4.9%	+4.7%	+4.2%	+4.3%	+5.2%	—	—	—
	100%	+0.8%	+4.0%	+3.3%	+3.0%	+2.7%	+2.4%	+2.9%	+3.0%	+3.7%

Median improvements over the best participant in the group when using collaborative BCIs for target localisation at difficulty level 5, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size							
		2	3	4	5	6	7	8	9
SC-cBCI	5%	-17.2%	-21.9%	—	—	—	—	—	—
	10%	-20.7%	-21.7%	-26.5%	-27.1%	—	—	—	—
	15%	-18.2%	-18.0%	-17.1%	-20.0%	-23.4%	-22.7%	—	—
	20%	-18.4%	-21.2%	-19.7%	-18.8%	-17.1%	-18.7%	-19.1%	—
	25%	-18.4%	-20.3%	-18.3%	-17.8%	-15.9%	-15.4%	-15.4%	-11.3%
	100%	-18.4%	-20.3%	-18.3%	-17.8%	-15.9%	-15.4%	-15.4%	-11.3%
MC-cBCI	5%	-14.3%	-12.6%	—	—	—	—	—	—
	10%	-13.7%	-15.7%	-15.5%	-18.9%	—	—	—	—
	15%	-12.4%	-13.0%	-14.8%	-15.7%	-16.7%	-16.8%	—	—
	20%	-12.1%	-9.9%	-10.5%	-9.1%	-8.4%	-8.0%	-7.3%	—
	25%	-11.6%	-9.6%	-9.6%	-8.1%	<i>-7.2%</i>	-7.1%	-5.0%	-6.5%
	100%	-11.6%	-9.6%	-9.6%	-8.1%	<i>-7.2%</i>	-7.1%	-5.0%	-6.5%
LDA-cBCI	5%	-14.0%	-7.1%	—	—	—	—	—	—
	10%	-11.6%	-11.7%	-8.8%	-11.3%	—	—	—	—
	15%	-9.4%	-8.5%	-9.0%	<i>-11.0%</i>	<i>-11.0%</i>	-11.8%	—	—
	20%	-6.2%	-3.4%	-3.2%	-1.5%	-1.4%	-2.0%	-0.4%	—
	25%	-5.8%	-3.2%	-2.0%	-0.1%	+0.5%	+2.8%	+2.9%	+2.1%
	100%	-5.8%	-3.2%	-2.0%	-0.1%	+0.5%	+2.8%	+2.9%	+2.1%

Median improvements over the best participant in the group when using collaborative BCIs for target localisation at difficulty level 6, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	-19.7%	-23.8%	-27.7%	-26.4%	-34.6%	—	—	—	—
	10%	-19.5%	-23.7%	-27.7%	-27.9%	-29.4%	-30.0%	—	—	—
	15%	-21.5%	-24.5%	-29.2%	-28.3%	-27.6%	-28.7%	-26.5%	—	—
	20%	-21.1%	-24.6%	-27.7%	-28.9%	-28.2%	-27.9%	-24.6%	-23.9%	—
	25%	-21.5%	-27.9%	-29.4%	-31.6%	-31.7%	-32.6%	-33.0%	-29.9%	-27.1%
	100%	-21.5%	-27.9%	-29.4%	-31.6%	-31.7%	-32.6%	-33.0%	-29.9%	-27.1%
MC-cBCI	5%	-12.5%	-17.4%	-16.6%	<i>-18.5%</i>	-19.3%	—	—	—	—
	10%	-13.6%	-16.5%	-17.6%	-18.1%	-19.0%	-18.7%	—	—	—
	15%	-13.7%	-14.5%	-15.2%	-14.8%	-14.5%	-14.4%	-13.8%	—	—
	20%	-13.5%	-13.7%	-13.8%	-13.7%	-13.5%	-13.0%	-12.4%	-11.6%	—
	25%	-13.7%	-15.1%	-15.0%	-15.0%	-14.5%	-14.4%	-14.0%	-15.1%	-14.0%
	100%	-13.7%	-15.1%	-15.0%	-15.0%	-14.5%	-14.4%	-14.0%	-15.1%	-14.0%
LDA-cBCI	5%	-10.3%	-18.0%	-16.3%	-17.9%	-18.0%	—	—	—	—
	10%	-11.8%	-16.2%	-16.4%	<i>-16.7%</i>	-17.8%	-17.6%	—	—	—
	15%	-12.4%	-13.1%	-13.7%	-14.3%	-14.1%	-13.8%	-13.5%	—	—
	20%	-11.9%	-12.3%	<i>-12.4%</i>	-12.9%	-12.8%	-12.5%	-11.9%	-11.9%	—
	25%	-12.3%	-12.6%	-12.7%	-12.8%	-12.8%	-12.7%	-12.7%	-12.4%	-13.4%
	100%	-12.3%	-12.6%	-12.7%	-12.8%	-12.8%	-12.7%	-12.7%	-12.4%	-13.4%

Median improvements over the best participant in the group when using collaborative BCIs for target localisation at difficulty level 7, as a function of group size and the dissimilarity-index threshold δ . Values in bold face are statistically significantly superior at the 1% confidence level according to a two-sample one-sided Kolmogorov-Smirnov test (group AUC vs maximum AUC of the group). Values in italics are statistically superior at the 5% confidence level.

Method	δ	Group size								
		2	3	4	5	6	7	8	9	10
SC-cBCI	5%	-20.9%	-21.5%	-17.6%	-15.0%	-17.0%	—	—	—	—
	10%	-16.8%	-17.8%	-18.6%	-18.0%	-17.9%	-14.0%	-18.9%	—	—
	15%	-16.6%	-15.7%	-16.8%	-16.1%	-17.8%	-14.0%	-18.9%	—	—
	20%	-17.5%	-17.1%	-15.8%	-16.2%	-15.9%	-16.3%	-18.9%	-19.3%	-26.0%
	25%	-17.5%	-17.1%	-15.8%	-16.2%	-15.9%	-16.3%	-18.9%	-19.3%	-26.0%
	100%	-17.5%	-17.1%	-15.8%	-16.2%	-15.9%	-16.3%	-18.9%	-19.3%	-26.0%
MC-cBCI	5%	-11.0%	-12.4%	-11.3%	-9.6%	-6.2%	—	—	—	—
	10%	-6.6%	-6.2%	-5.8%	-5.0%	-5.6%	-4.4%	-4.0%	—	—
	15%	-5.3%	-3.8%	-3.6%	-3.9%	-5.2%	-4.4%	-4.0%	—	—
	20%	-4.6%	-2.1%	-1.4%	+0.6%	+2.0%	+2.7%	+4.0%	+4.6%	+5.5%
	25%	-4.6%	-2.1%	-1.4%	+0.6%	+2.0%	+2.7%	+4.0%	+4.6%	+5.5%
	100%	-4.6%	-2.1%	-1.4%	+0.6%	+2.0%	+2.7%	+4.0%	+4.6%	+5.5%
LDA-cBCI	5%	-12.8%	-14.2%	-12.8%	-11.6%	-9.2%	—	—	—	—
	10%	-8.1%	-8.2%	-7.2%	-6.8%	-7.2%	-7.4%	-8.6%	—	—
	15%	-3.4%	-4.2%	-4.7%	-6.1%	-7.2%	-7.4%	-8.6%	—	—
	20%	-3.1%	-2.9%	-1.5%	-0.5%	+2.3%	+3.8%	+3.9%	+3.5%	+1.7%
	25%	-3.1%	-2.9%	-1.5%	-0.5%	+2.3%	+3.8%	+3.9%	+3.5%	+1.7%
	100%	-3.1%	-2.9%	-1.5%	-0.5%	+2.3%	+3.8%	+3.9%	+3.5%	+1.7%

Appendix D

Individual ERD/ERS Analysis in Response to Cuts

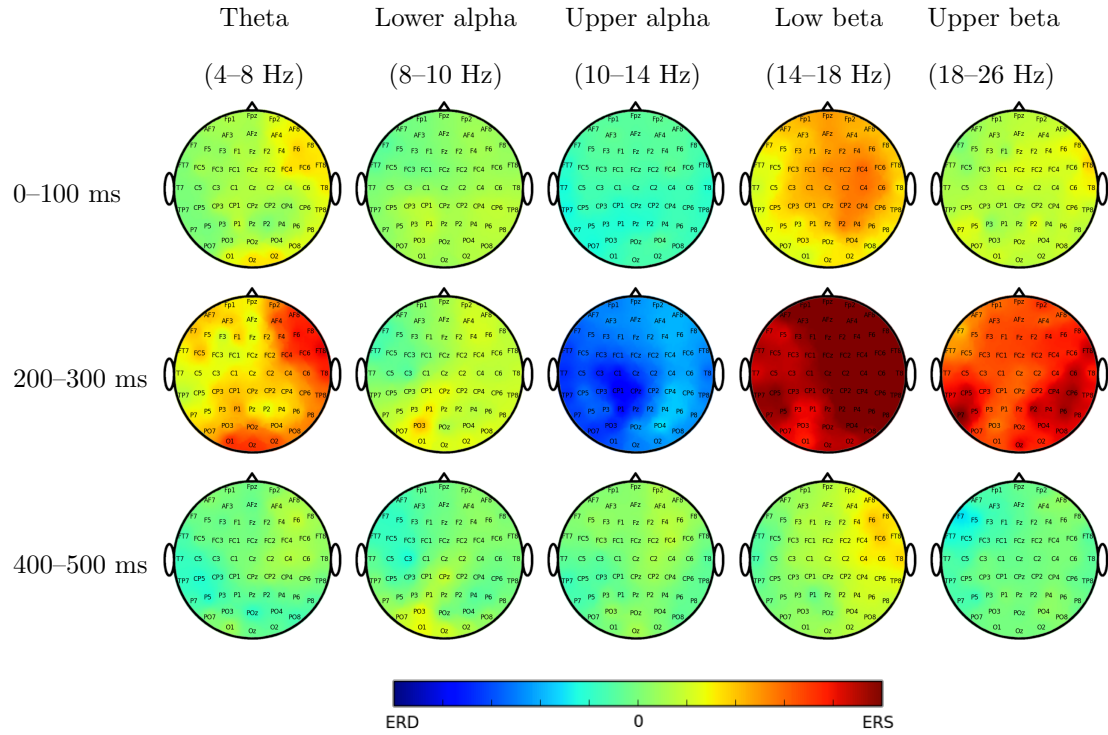
This appendix contains additional results for the frequency analysis sections from Chapter 7.

In particular, it contains a section devoted to illustrating the differences in individual ERD/ERS patterns as opposed to the grand average that was provided in the chapter, and a grand average of coherence in the pre- and post-cut conditions to which no threshold has been applied, so the magnitude of all connections can be seen.

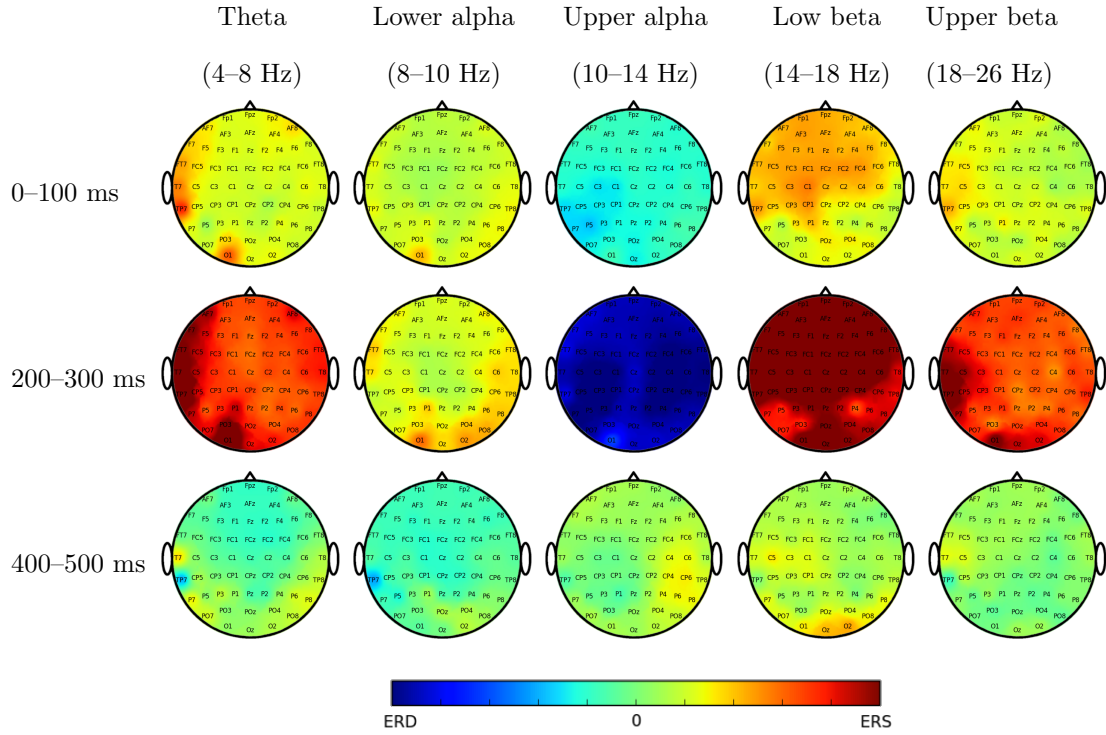
D.1 Individual ERD/ERS Topoplots

This section contains the ERD/ERS analysis from Chapter 7, which has been performed on an individual-by-individual basis, across all movies.

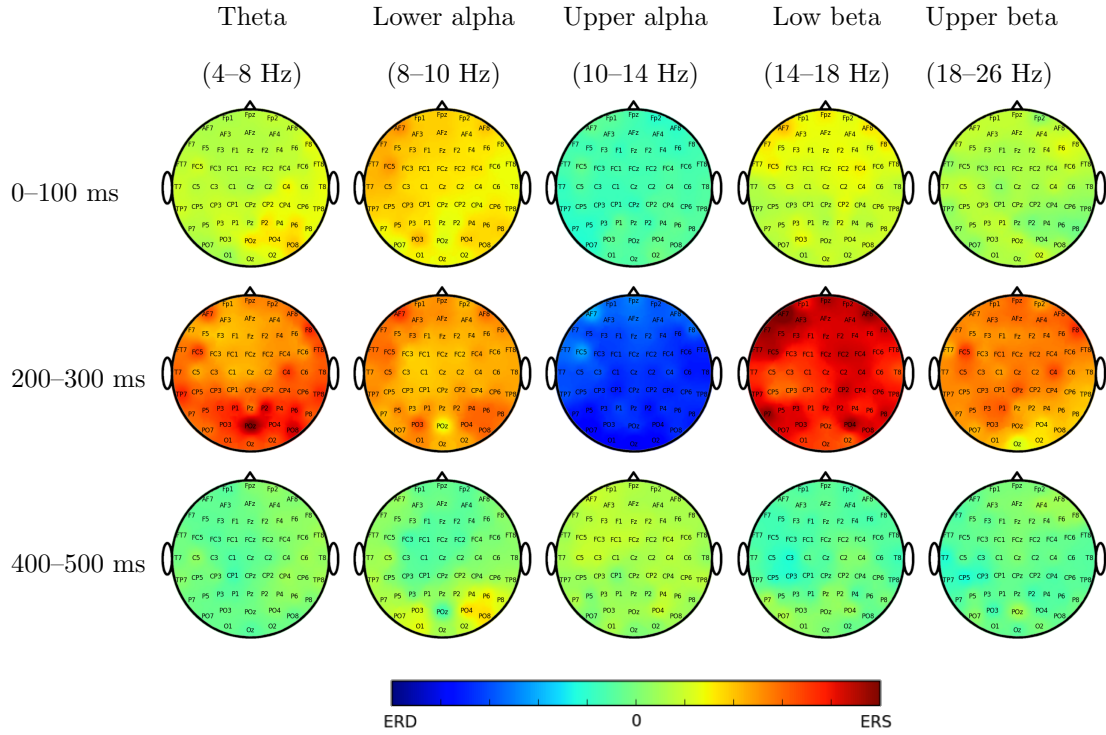
In order to make comparisons easier, the color scale has been kept constant across all participants



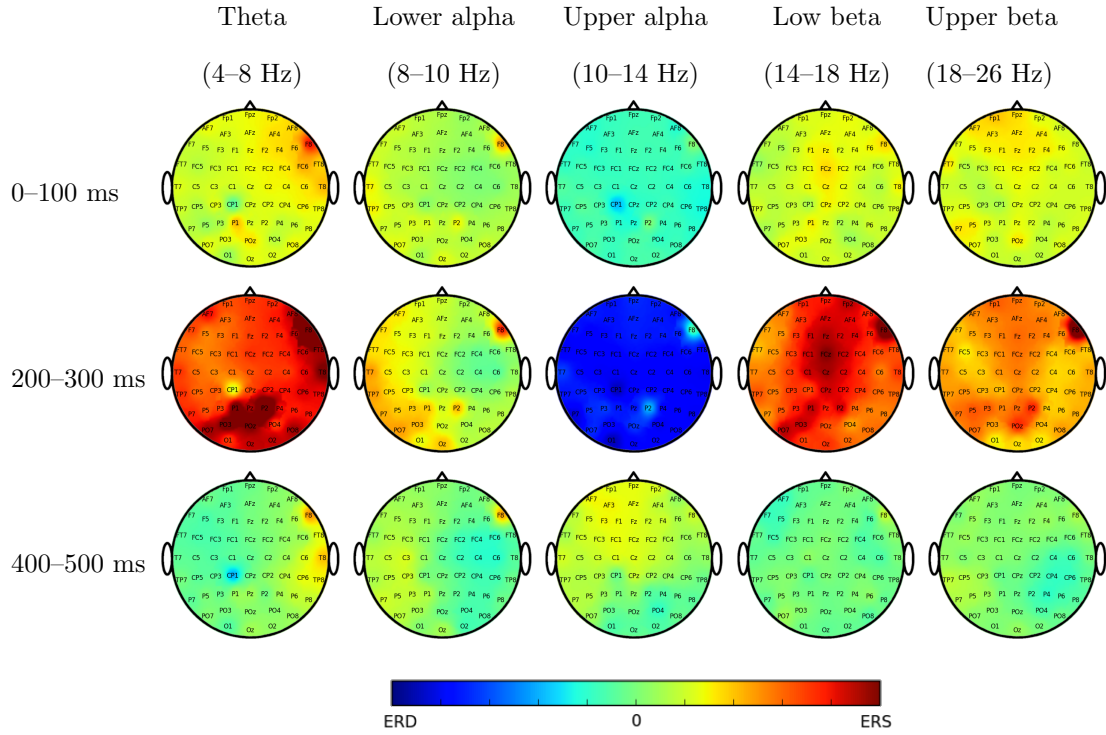
Spatial distribution of ERD/ERS after the occurrence of a cut for participant 001 at different time intervals and frequency bands.



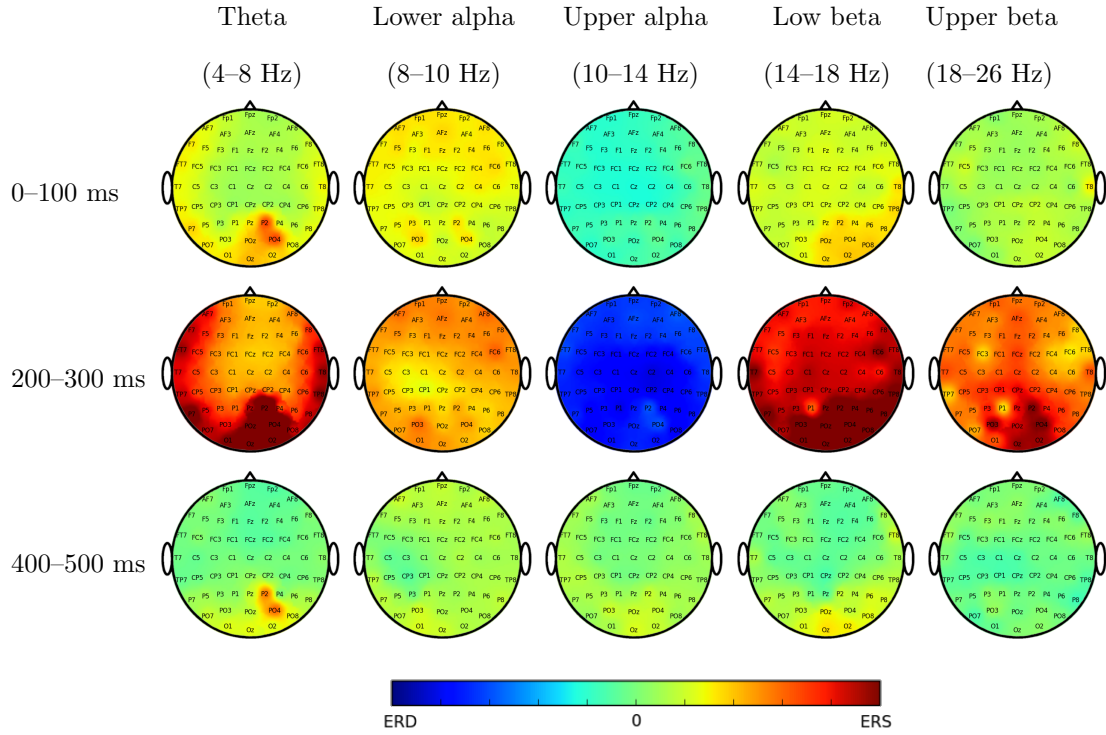
Spatial distribution of ERD/ERS after the occurrence of a cut for participant 001 at different time intervals and frequency bands.



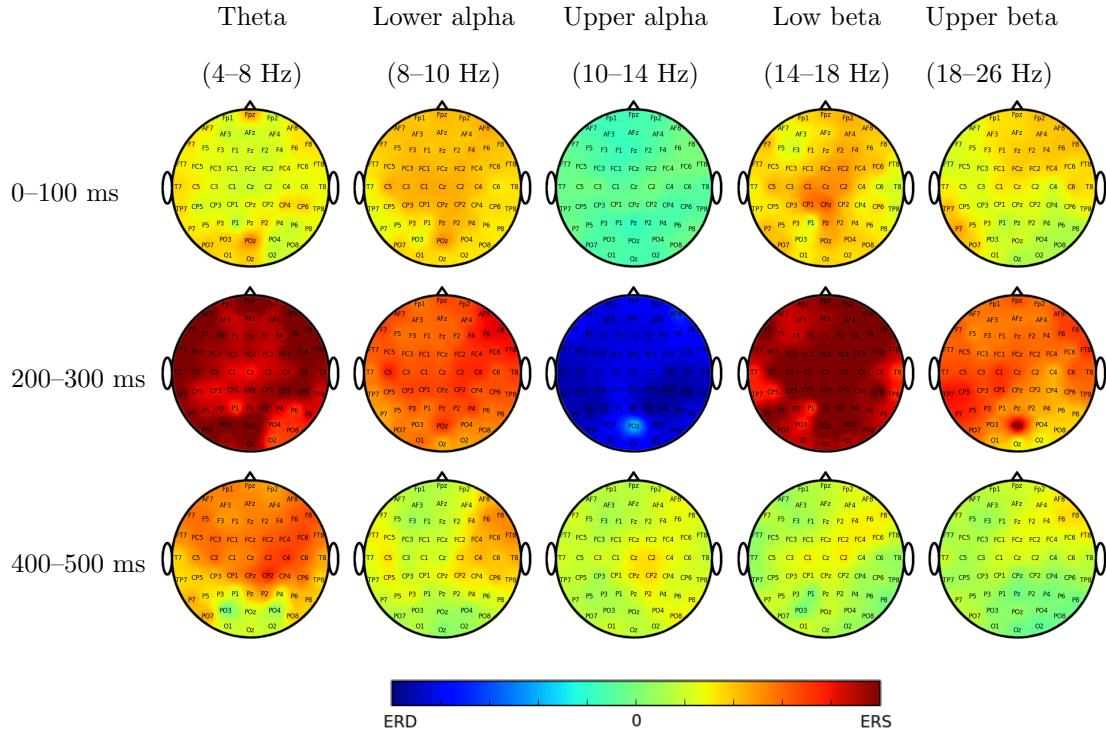
Spatial distribution of ERD/ERS after the occurrence of a cut for participant 001 at different time intervals and frequency bands.



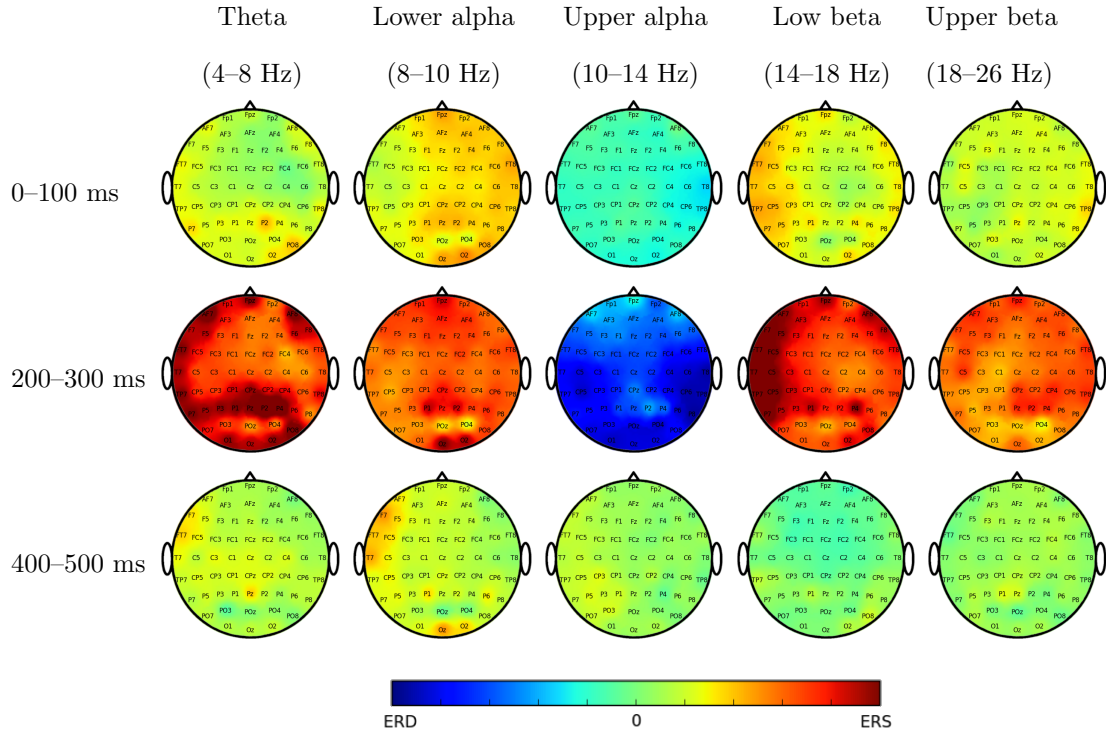
Spatial distribution of ERD/ERS after the occurrence of a cut for participant 001 at different time intervals and frequency bands.



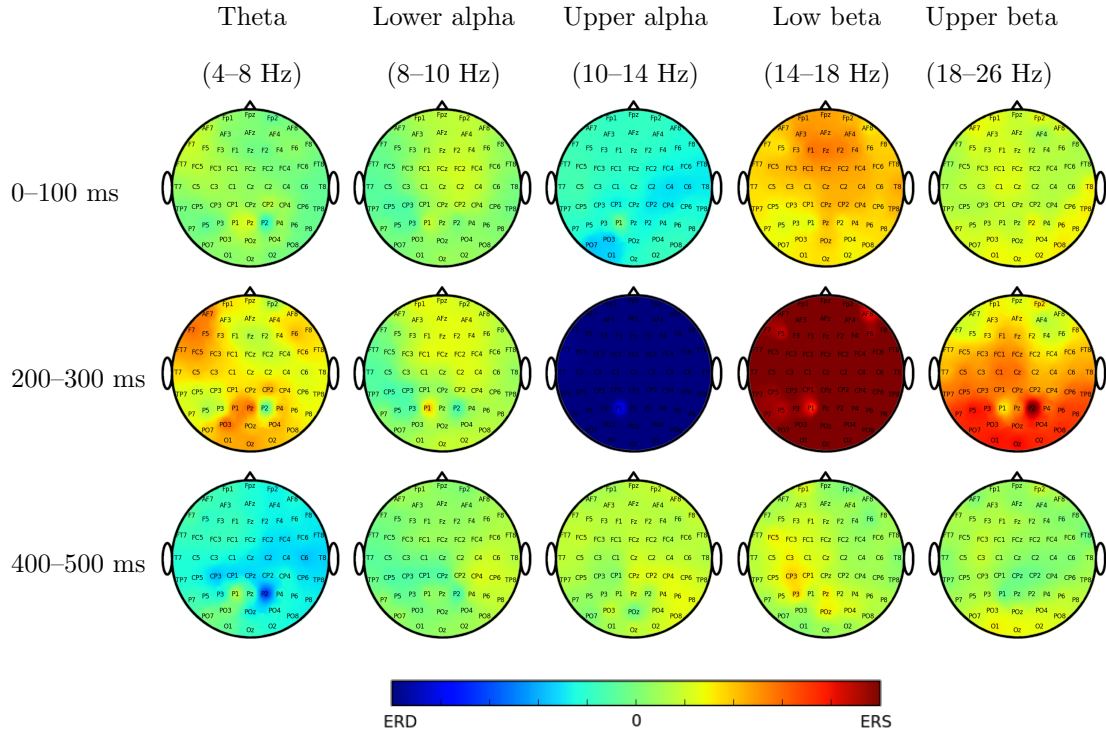
Spatial distribution of ERD/ERS after the occurrence of a cut for participant 001 at different time intervals and frequency bands.



Spatial distribution of ERD/ERS after the occurrence of a cut for participant 001 at different time intervals and frequency bands.

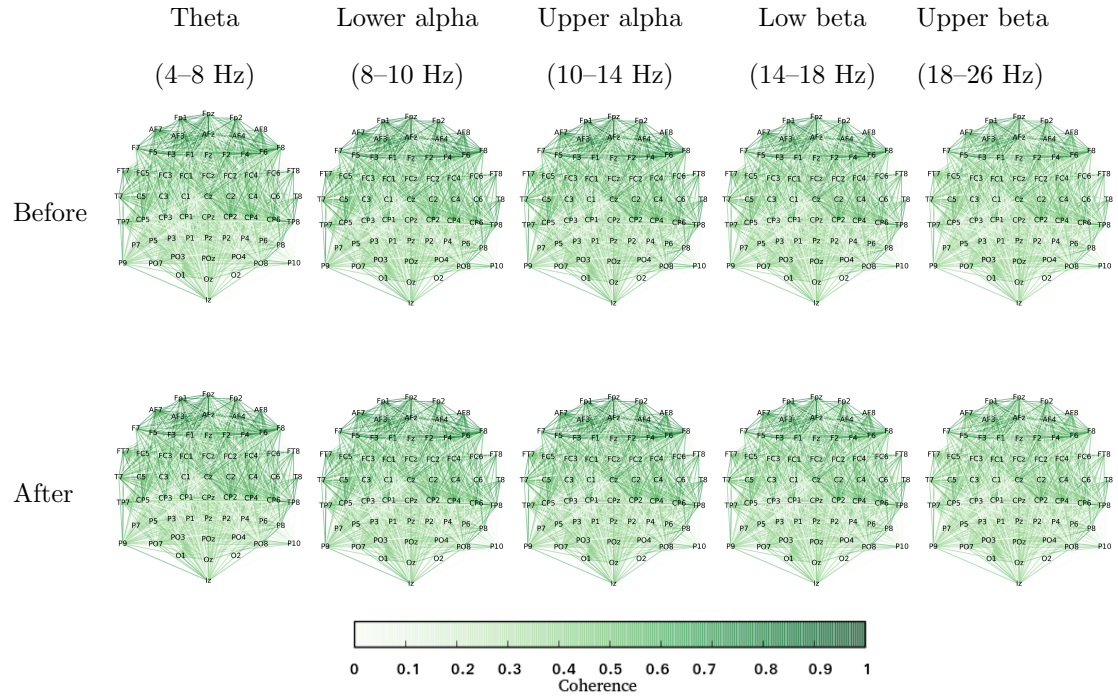


Spatial distribution of ERD/ERS after the occurrence of a cut for participant 001 at different time intervals and frequency bands.



Spatial distribution of ERD/ERS after the occurrence of a cut for participant 001 at different time intervals and frequency bands.

D.2 Grand Average Coherence Plots Preceding and Following Cuts



Average coherence between pairs of electrode regions before (top row) and after (bottom row) the occurrence of a cut in different frequency bands.